

THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

Verification de la precision d'un capteur de tracteur agricole et description d'un algorithme de correction

Maillard, Damien

Award date:
2021

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Université de Namur
Faculté d'informatique
Année académique 2020-2021
Literature review

Vérification de la précision d'un capteur de tracteur agricole et description d'un algorithme de correction.

Damien Maillard



Maître de stage :

Promoteur : *M^{me} Marie-Ange Remiche* (Signature pour approbation du dépôt- REE art. 40)

Co-promoteur : *M^r Benoit Frenay*

Confidentialité du mémoire : Mémoire présenté en vue de l'obtention du grade de Master en Sciences Informatiques.

Remerciements

Il est important de commencer ce mémoire par un remerciement au corps professoral de l'Université de Namur pour leur apport de connaissances et l'ouverture à l'esprit critique durant ma formation.

Tout d'abord, je voudrais remercier Madame Marie-Ange Remiche ainsi que Monsieur Benoit Frenay pour leurs conseils lors de la rédaction de ce mémoire.

Merci également de m'avoir permis de travailler sur ce thème d'analyse.

Je voudrais également remercier, de manière générale, mes collègues du Centre de Recherches Agronomiques Wallon pour leur patience et leurs conseils.

Enfin, je voudrais remercier ma famille pour leur soutien et leur patience lors de mes longues soirées de travail.

Résumé

L'agriculture de précision est de plus en plus utilisée dans les exploitations agricoles. Cette agriculture de précision se base sur la collecte de données de sol, de données environnementales ou de données d'utilisation de machines agricoles pour élaborer des calculs de production et essayer d'en améliorer les rendements.

Ces collectes de données sont souvent importantes et les systèmes informatiques doivent être dimensionnés pour permettre cela. Plusieurs notions doivent donc être prises en compte afin d'en faciliter tant la collecte des données que l'exploitation de celles-ci.

Il faut que ces données soient :

- collectées correctement ;
- fiables ;
- stockées de manière efficace ;
- utilisées à bon escient.

La question est donc posée de savoir si, lors de ces collectes, les données obtenues sont fiables. Quel peut être le taux d'erreur d'une donnée fournie et quel sera l'impact sur les résultats des calculs proposés utilisant ces données.

Cette question peut également être étendue à l'impact sociétal ou environnemental de ces résultats.

Le choix s'est posé d'essayer de répondre à ces questions en consultant les données collectées sur un tracteur agricole.

Un récapitulatif des études menées à ce sujet est tout d'abord proposé.

Plusieurs études mettent l'accent sur le fait que les données fournies par les capteurs d'un tracteur agricole ne peuvent être exploitées sans application d'une méthode de régression.

Ensuite, un essai de vérification de ces affirmations est réalisé par l'utilisation de jeux de données collectées sur un tracteur agricole suivant des conditions d'utilisation déterminées.

Enfin, pour terminer, un aperçu des études menées qui proposent des algorithmes permettant d'appliquer les corrections nécessaires est rédigé. La correction peut être proposée via une méthode de régression ou par le biais de réseaux de neurones. Une proposition consistera à implémenter ultérieurement les algorithmes analysés pour permettre aux chercheurs de valider des données qui serviront par la suite dans des modèles mathématiques.

Glossaire

ASABE :	American Society of Agricultural and Biological Engineers. Organisation scientifique et éducative américaine destinée à l'évolution des techniques d'ingénierie dans le domaine de l'agriculture, de l'alimentation et des systèmes biologiques.
AUTOSAR :	(AUTomotive Open System ARchitecture) Architecture logicielle standardisée et ouverte pour les unités de contrôle électroniques ECU qui fournit un ensemble de spécifications pour un format d'échange standardisé.
BPNN :	Back Propagation Neuronal Network. Est un réseau de neurones fonctionnant sur base d'un mécanisme de rétro-propagation. C'est grâce à ce mécanisme que les réseaux de neurones apprennent.
Bus-CAN :	Le bus de données CAN (Controller Area Network) est un système de transmission de données et de communication série. Il met en application une connexion d'éléments par la méthode de multiplexage. Il permet d'établir des priorités entre les entités communicantes.
BSFC :	Brake Specific Fuel Consumption. Consommation de carburant spécifique au freinage. Il peut être considéré comme la consommation de carburant spécifique à la puissance du moteur.
CCP :	CAN Calibration Protocol est un protocole maître-esclave basé sur CAN qui est utile dans la calibration et l'acquisition des données transitant sur le Bus-CAN.
DGPS :	Differential Global Positioning System Le système de positionnement global différentiel est une amélioration du GPS classique permettant une meilleure

	précision. Il utilise un réseau de stations fixes de référence.
ECU :	Electronic Control Unit. Système électronique embarqué qui commande des dispositifs physiques au sein d'un moteur de véhicule (voiture, camion, bus, tracteur...).
GPS-RTK :	GPS-RTK (Real Time Kinematic) peut se définir comme étant une technique de positionnement satellite (Global Positioning System). Le système réceptionne des corrections d'erreurs fournies par une station de référence. Ces corrections sont fournies en temps réel. Le GPS-RTK offre une meilleure précision que le DGPS.
Haversine :	La méthode Haversine, inventée au début du 19 ^{ème} siècle, permet de déterminer une distance entre deux points d'une sphère par le biais des coordonnées latitude et longitude de chacun des points. C'est un cas spécifique d'une formule de trigonométrie sphérique dérivée du Sinus Verse.
ISOBUS :	ISOBUS est un protocole de communication basé sur la norme J1939 et destiné à l'industrie agricole. Il permet la communication d'un tracteur agricole avec les engins agricoles (outils) qui lui sont attelés.
LTFT :	Long Term Fuel Trim. C'est le contrôle du temps d'ouverture de l'impulsion d'injection de carburant dans le moteur, sur toute la plage de fonctionnement de celui-ci.
NI-TDMS :	Solution de gestion de données techniques (Technical Data Management Solution) créée par la société National Instruments.
OBD-II :	On Board Diagnostics version 2. OBD offre la capacité de fournir un diagnostic et un rapport d'un véhicule, permettant à l'opérateur du véhicule ou technicien de connaître l'état des sous-systèmes du moteur.
RBF :	Réseau de neurones RBF (Radial Basis Function). Une fonction à base radiale est une fonction dont la valeur ne dépend que de la distance à l'origine. Elle se définit elle-même par la distance entre l'origine et son centre.
SAE J1939 :	Le SAE J1939 est un protocole de communication qui est utilisé dans le bus de communication Bus-CAN.
SFC :	Specific Fuel Consumption. Spécification technique utilisée pour quantifier le rendement énergétique d'un moteur en

	fonction de sa puissance.
STFT :	Short-Time Fourier Transform. La "transformée de Fourier" est une transformation permettant de déterminer la fréquence d'une sinusoïdale.
SVM :	Support Vector Machine. Machine à vecteurs de support. Ensemble de techniques d'apprentissage supervisé destiné à résoudre des problèmes de régression.

Table des matières

1	Introduction	9
1.1	Contexte	9
1.2	Finalité de l'étude	10
1.3	Un projet actuellement en cours d'étude au CRA-W	11
1.4	Quelques notions introductives	12
2	Fiabilité des capteurs - Revue	14
2.1	Consommation de carburant	14
2.2	Puissance moteur	19
2.3	Patinage des pneus	20
2.4	Couple et régime moteur	21
2.5	ECU et Etalonnage	24
3	Analyse de données disponibles	27
3.1	Fiabilité de positionnement GPS	29
3.2	Hypothèses de départ	33
3.3	Choix des données pour cette première analyse	34
3.4	Choix des fichiers de données	35
4	Méthodologie de l'analyse des données	38
4.1	Collecte des données.	38
4.1.1	Nettoyage du fichier contenant les données	39
4.2	Calcul de la vitesse	40
4.2.1	Script R de calcul de la vitesse	41
4.2.2	Vérification des vitesses calculées pour le premier relevé	44
4.2.3	Vérification des vitesses calculées pour le deuxième relevé	52
4.3	Comparaison de la vitesse avec le nombre de tours/minutes . .	56
4.3.1	Premier relevé avec mesures du capteur GPS-2	56

4.3.2	Deuxième relevé avec mesures du capteur GPS-1	57
5	Algorithmes de correction	59
5.1	Réseau de neurones	61
5.2	Machine à vecteurs de support	67
6	Conclusion	69

Chapitre 1

Introduction

Est-il possible de déterminer si un capteur de tracteur agricole est fiable ?

1.1 Contexte

De plus en plus, l'approche de l'agriculture de précision s'installe dans les pratiques d'exploitation des cultures agricoles [1]. Cette agriculture de précision tend à s'appuyer sur une analyse des éléments constituant l'environnement pour affiner au mieux les pratiques agricoles et adapter les réactions utiles en fonction des scénarios environnementaux.

Par exemple, sur base d'une analyse fine concernant les prévisions météorologiques d'une zone agricole exploitée, on peut arriver à définir avec exactitude les risques de maladies, de gel, etc.

Ces nouvelles méthodes de collecte des données se basent, pour l'essentiel, sur l'utilisation de capteurs qui mesurent des informations concernant l'environnement d'utilisation. Ces informations sont stockées dans des centres de données et peuvent ensuite être utilisées dans les modèles mathématiques de prédiction. Le capteur devient donc le coeur du mécanisme global, et on peut donc légitimement se poser la question de sa fiabilité.

Les informations fournies par les capteurs, peuvent vite prendre une grande importance dans la chaîne de traitement de calculs prévisionnels [2]. Plusieurs enjeux se font face pour améliorer cette chaîne de traitement : il s'agit de l'interopérabilité entre les éléments agricoles (tracteurs, outils agricole, etc) pour permettre la circulation des données des capteurs, mais également de la normalisation des formats de données, de l'élaboration des protocoles d'échange de données, du dialogue entre les différents acteurs producteurs de machines agricoles, etc.

L'importance, quant à elle, est liée au stockage, à la valeur des données (fiabilité et exactitude), à leur accès et à leur redondance.

Cette présente analyse synthétise les études déjà réalisées qui tentent de répondre à la question de la validité des données.

1.2 Finalité de l'étude

L'objectif premier de cette étude est d'essayer de vérifier si les capteurs utilisés pour favoriser l'émergence de l'agriculture de précision sont fiables et s'ils peuvent fournir des informations de base utilisables dans les modèles de calculs. Il est également intéressant de vérifier si des études font état de mécanismes de correction des erreurs dans les jeux de données fournies par un capteur.

A une échelle plus large, on pourrait estimer quel est l'impact sociétal et environnemental des éventuelles erreurs produites par les capteurs dans leur utilisation agricole. Cette analyse n'aura néanmoins pas la prétention de faire une analyse de ceux-ci mais les conclusions pourront servir de base pour une recherche future.

Dans le cadre de cette présente analyse, la question se focalise sur la donnée fournie par le capteur d'un tracteur agricole. Elle tentera de vérifier les résultats des études présentées par la consultation de jeux de données à notre disposition et d'y détecter les éventuelles anomalies. Ensuite, plusieurs propositions seront faites pour déterminer s'il est possible d'effectuer des corrections sur ces données erronées et vérifier si ces corrections apportent

des modifications importantes dans les résultats.

Puisqu’une donnée va avoir un cycle de vie relativement long dans un processus informatisé d’une exploitation agricole, il peut être opportun de vérifier cette précision.

La chaîne de processus décisionnel d’une exploitation agricole, basée sur un traitement informatisé, peut être définie comme étant plusieurs types d’applications distinctes (souvent appelé OAD – outil d’aide à la décision). Elle permet de :

- définir les rendements de production d’une exploitation ;
- calculer les émissions de GES (Gaz à Effet de Serre) d’une exploitation ;
- d’améliorer les processus de productions, tant agricoles (culture et travail de champ, pulvérisation, semis, stockage) que d’élevage ;
- de mieux maîtriser les coûts de matières premières ou de consommation d’énergies ;
- ...

Les véhicules agricoles actuels sont équipés d’un ensemble de capteurs permettant de mesurer les comportements des divers éléments du tracteur. Ces capteurs sont reliés à un faisceau de connexion appelé Bus-CAN [3] qui est régi par une norme (un rappel de la norme est décrit plus bas).

Il est donc aisé de capturer les données fournies par chaque capteur et de les stocker numériquement pour une consultation ultérieure.

1.3 Un projet actuellement en cours d’étude au CRA-W

Le projet nommé « GéoCAN » (<https://www.cra.wallonie.be/fr/geocan>) prévoit d’enregistrer les comportements de tracteurs agricoles en cours d’utilisation pour permettre d’élaborer des résultats qui aideront à la décision d’amélioration de ces comportements.

Ce projet prévoit d'utiliser les informations fournies par les différents capteurs qui équipent le tracteur agricole. Les informations sont collectées lorsque ce tracteur est en cours de travail agricole.

Il est encore en cours de développement et des recherches doivent également être menées pour définir les modèles mathématiques de calculs en fonction des résultats souhaités.

La donnée du tracteur devra :

- être collectée ;
- être transformée pour devenir lisible dans un système informatique standard (donc non dédié) ;
- être acheminée vers le lieu de stockage ;
- être précise ;
- ne pas être dupliquée ;
- son ensemble devra contenir assez de données sans être trop volumineux.

La finalité de l'étude sera de proposer une réflexion sur l'intérêt d'insérer un algorithme complémentaire dans l'application GéoCAN. Cet algorithme aura pour but d'effectuer une correction de la précision des données si besoin. Il pourra être utilisé dans l'analyse pour créer des données destinées aux modèles mathématiques actuellement en cours de recherche.

1.4 Quelques notions introductives

Pour interpréter correctement les données des capteurs, il est nécessaire de comprendre le fonctionnement de ces capteurs et des connexions qui sont établies entre eux. Une brève explication, nécessaire à la compréhension des différentes notions utilisées, est décrite ci-dessous.

Bus-CAN :

Comme pour tout autre véhicule motorisé, les tracteurs agricoles sont équipés de systèmes de capteurs reliés entre eux par un système de connectivité. Ce système globalisé prend le nom de Bus-CAN (Controller Area Network) et est normalisé via la norme ISO-11898 [2 - 4].

Le Bus-CAN fournit des données d'utilisation du tracteur agricole (nombre de tours/minutes du moteur, vitesse enclenchée, coordonnées GPS, état d'utilisation des éclairages, vitesse enclenchée dans la boîte de vitesse, etc).

La norme Bus-CAN se base sur le multiplexage (technique consistant à faire passer plusieurs informations distinctes via un seul canal de transmission).

SAE J1939 :

SAE J1939 [2 - 5] est un protocole de communication utilisé par les capteurs connectés au Bus-CAN. Il décrit l'utilisation des diverses couches du modèle OSI.

Ce protocole de communication décrit la structure d'une trame de données de manière à ce que celle-ci soit lisible par l'ensemble des éléments connectés au bus de communication.

Il décrit notamment le nombre de bits réservé à la priorité, aux données, à l'adresse source, au type d'envoi (point à point ou broadcast), etc.

Les données d'un capteur sont identifiées par deux caractéristiques distinctes :

- **Parameter Group Number (PGN)** : identifiant unique d'une trame de données à l'intérieur du standard SAE J1939. Le numéro PGN peut rassembler plusieurs capteurs sous le même numéro ;
- **Suspect Parameter Number (SPN)** : identifiant unique d'un signal contenant des données. Il peut être l'identifiant de la donnée fournie par un capteur. Les SPN sont regroupés par PGN et sont décrits en terme de bit de début, de bit de longueur...

ISOBUS :

ISOBUS [2 - 6] est une extension de la norme SAE J1939 spécifique pour les engins agricoles. Cette norme permet notamment de faire dialoguer les données spécifiques d'un tracteur avec les données des machines agricoles (outils) complémentaires qui sont attelées à un tracteur pour le travail au champ. Elle est régie par la norme ISO 11783 [2 - 7].

Chapitre 2

Fiabilité des capteurs - Revue

La question de la fiabilité des données est présente dans plusieurs études. Aussi, par souci de lisibilité, les études décrites ci-dessous sont classées par type de capteur analysé ou par type d'usage.

La première finalité de l'utilisation des données Bus-CAN d'un tracteur en travail agricole est de réduire sa consommation de carburant. Les prix du carburant devenant de plus en plus élevés, cette problématique de réduction de consommation devient un enjeu majeur des exploitants agricoles pour leur permettre de garder des marges bénéficiaires valables.

De plus, la réduction de la consommation de carburant est également un enjeu global concernant la production de GES (Gaz à Effet de Serre). La recherche se concentre donc sur ces aspects.

2.1 Consommation de carburant

Dans le Bus-CAN, la valeur fournie de consommation de carburant est une valeur calculée au départ de plusieurs valeurs atomiques.

Deux méthodes voient le jour pour établir une consommation moyenne de carburant. Ces deux méthodes se basent sur des relevés d'utilisation de plus de 500 tracteurs et utilisent le régime du moteur (nbrTours/minutes en rpm) de chacun de ces relevés pour estimer la consommation. La méthode générale (première des deux méthodes) sera, plus tard, adoptée en tant que

norme ASABE (2011).

Néanmoins, beaucoup de paramètres ne sont pas pris en compte dans cette norme. Une étude [8] se penche donc sur la consommation de carburant par le biais de relevés de données fournies par le Bus-CAN. Il est constaté que les variations de carburant sont observables à la fin d'une opération en ligne droite de travail du sol et, également, lors d'une opération de demi-tour du tracteur. Il existe donc une relation directe entre la consommation de carburant et le type d'opération en cours. Il est donc nécessaire de sélectionner les plages de valeurs pour lesquelles le moteur a une rotation en régime permanent.

Le régime moteur variant en permanence dans une plage de 1900 tr/min à 2100 tr/min, cette variation entraînera, malgré tout, des erreurs d'estimation. Le choix est donc fait de calculer la puissance de sortie du tracteur en fonction de la puissance nominale de la prise de force, de la résistance au mouvement, de l'efficacité de traction et de l'efficacité énergétique, le tout combiné avec la vitesse du tracteur (EngineSpeed). Le résultat permet de calculer le taux de consommation par type de travail (travail de champ, demi-tour, etc).

Beaucoup de paramètres ne sont toujours pas pris en compte ou estimés via les valeurs moyennes obtenues sur base de précédents relevés. L'utilisation d'une formule de régression sera utilisée pour calibrer le taux de consommation, en y retirant les données pour les plages de valeurs ne correspondant pas à une période de travail (moteur tournant au ralenti et transitions).

S'il reste difficile de représenter la valeur instantanée, le modèle proposé dans cette étude est meilleur que le modèle ASABE (2011).

La question de la fiabilité des données fournies par le Bus-CAN est posée dans l'étude de Samuel E. Marx, 2015 [9]. Elle propose de comparer les valeurs de consommation mesurées via le Bus-CAN avec des valeurs relevées via des appareils de mesure de consommation instantanée. De par le fonctionnement intrinsèque du système d'injection de carburant dans le moteur, il est possible d'obtenir des valeurs différentes entre la mesure du débit de carburant calculée via le Bus-CAN et la mesure effectuée physiquement.

Etant donné que la consommation fournie par le Bus-CAN est calculée, des problèmes de synchronisation peuvent dégrader la transmission des données. Ces problèmes de synchronisation peuvent également dégrader la précision des données servant aux calculs.

Il faut également tenir compte de l'acquisition et du volume de stockage des données. Trois méthodes d'acquisition et de stockage sont proposées et analysées pour vérifier si les valeurs se dégradent dans l'une ou l'autre méthode. La question de fiabilité devient donc légitime dès le moment où un grand nombre de données est disponible. Ce grand nombre de données pose le défi du stockage, de son accès une fois stocké, mais également de la synchronisation avant stockage.

Une fréquence d'échantillonnage élevée ($100 \text{ hz} = 100$ fois par seconde) modifie fortement la taille du fichier à collecter. Etant donné que la capacité de stockage physique n'est pas infinie, plus la fréquence d'échantillonnage augmente et plus l'espace de stockage sera important.

Deux méthodes ont été utilisées pour enregistrer les données à des fréquences d'échantillonnage différentes. Une troisième méthode consiste à enregistrer les données sous forme d'ondes à une fréquence de 2 hz (2 fois par seconde) en valeurs pondérées. Le choix d'utiliser des valeurs pondérées résulte d'une volonté d'obtenir des fichiers de stockage moins volumineux.

L'analyse des trois méthodes distinctes d'enregistrement ne permet pas de définir que les valeurs relevées sont moins précises par le biais de l'une ou l'autre méthode. Une première conclusion de l'étude permet de constater que la capture des données, quelle que soit la méthode utilisée, ne favorise pas l'émergence d'un risque d'erreur dans la précision des données.

Cette étude propose une analyse en laboratoire de test. Plusieurs tracteurs sont sélectionnés pour ce test et ils sont soumis à des régimes moteurs différents. Le débit de carburant mesuré par le biais des données J1939 comporte une erreur de $\pm 5\%$ en rapport au débit de carburant réel mesuré physiquement. Mais lorsque le couple moteur augmente, l'erreur peut redescendre dans une estimation de $\pm 1\%$. Plusieurs charges moteur distinctes ont également été sélectionnées pour effectuer les vérifications

nécessaires. Néanmoins, si ces tests pouvaient être reproduits en activité de travail agricole du tracteur, l'erreur de précision serait certainement plus importante de par le fait d'une variation constante de régime ou de charge du moteur.

L'étude de S. Lacour et al., 2014 [10] cherche à améliorer les mesures de consommation de carburant d'un tracteur. Elle établit que les calculs de consommation de carburant, basés sur les valeurs moyennes annuelles, montrent leurs limites, les utilisateurs demandant des bilans carburants plus détaillés.

Un modèle est conçu pour permettre de calculer les valeurs de consommation de carburant en fonction des différents types de travail de champ. Il convertit une force de traction en une consommation de carburant et est calibré en fonction de mesures faites en laboratoire sur banc d'essai. Les différentes opérations de travail agricole sont découpées en sous-tâches de manière à pouvoir les catégoriser correctement et, au besoin, en scindant certaines dans l'analyse des résultats. Les opérations scindées sont celles qui concernent le fonctionnement au ralenti, le demi-tour du tracteur en fin de ligne, etc.

Il est à signaler que le fait d'utiliser les données du tracteur en situation réelle de travail agricole permettra d'obtenir un grand volume de données à faible coût par rapport à l'utilisation d'un banc d'essai.

Les dégradations de l'environnement liées à l'utilisation de moteur à combustion se vérifient tant dans les zones rurales que dans les zones urbaines par l'utilisation de véhicules de transport agricole [11]. Pour réduire l'impact sur l'environnement de la consommation de carburant, cette étude a mesuré le rendement maximal d'un véhicule agricole. Il s'agit de vérifier à quelle vitesse du moteur la consommation est la moins élevée. Les mesures ont été réalisées pour une plage de vitesse de rotation allant de 1600 tr/min à 2200 tr/min. Ce rendement maximal est atteint à une vitesse de 1750 tr/min. Le moteur a une valeur maximale de son efficacité lorsque la vitesse de rotation est inférieure à la vitesse nominale et supérieure au couple maximal.

La question de l'impact de la consommation de carburant en fonction du

comportement d'utilisation du tracteur est posée [12]. Lorsque le conducteur appuie sur la pédale d'accélération, l'admission d'air dans le moteur change. Le rapport air-carburant intervient directement sur les performances et la puissance d'un moteur, les émissions (Oxydes d'Azote, Monoxyde de Carbone) produites et la consommation. L'injection de carburant est contrôlée par l'ECU. Des capteurs sont utilisés pour mesurer le débit d'air, puis envoyer l'impulsion correcte à l'ECU. Le capteur de compensation de carburant de type STFT (Short-Time Fourier Transform) permet de corriger de manière instantanée le mélange air-carburant envoyé dans l'injection. Ce capteur est lié aux changements immédiats de valeur d'admission d'air (se produisant plusieurs fois par seconde) alors que le capteur de type LTFT (Long Term Fuel Trim) permet des changements moyens au fil du temps.

Dans cette étude, réalisée pour des véhicules sur route, la charge signifie la quantité d'énergie requise par le moteur pour que le véhicule ait de la vitesse et de l'accélération. Lorsqu'un véhicule roule sur l'autoroute, il a besoin d'un petit pourcentage de sa puissance de sortie totale disponible pour maintenir la vitesse. La charge peut être calculée en fonction de la formule suivante :

$$EngineLoad = \frac{CurrentAirflow}{MaxAirflow(Rpm) \times \frac{Baro}{29.92} \times \sqrt{\frac{298}{Tamb+273}}} \quad (2.1)$$

Avec :

- Current Airflow est la mesure de débit d'air dans le moteur ;
- Max Airflow est la capacité maximale d'air pour le moteur ;
- Baro est la pression barométrique mesurée par le capteur d'air ;
- Tamb est la température ambiante qui est stable.

La charge est le principal facteur qui détermine le débit de carburant et donc la consommation. Celle-ci change lorsque la vitesse augmente, ce qui signifie qu'elle est principalement affectée par le type de conduite (accélération, décélération).

Il en est déduit que les phénomènes de rotation d'un tracteur (demi-tour et changement de direction), d'accélération après rotation ou des contraintes liées au sol (changement de type de sol, changement de pente de terrain), auront un impact important sur la consommation de carburant.

Plus la vitesse est constante (sans accélération ou décélération) et moins la quantité de carburant instantanée est utilisée. A l'inverse, plus la consommation instantanée de carburant varie et plus le risque d'erreurs est élevé. Cette étude est prévue pour des véhicules autres que des tracteurs agricoles mais il est supposé qu'il en est de même pour les tracteurs.

2.2 Puissance moteur

L'amélioration des processus de rendement agricole fait partie des enjeux à étudier. Pour cela, il est nécessaire de définir correctement la charge utile que le tracteur devra être capable de supporter [13].

Un tracteur agricole en travail de champ est soumis à de nombreuses contraintes. Le comportement du tracteur en travail variera en fonction des conditions de sol (terrain sec, ornières, etc) et des conditions environnementales (pluie, humidité, température, etc). De plus, la répartition de la force du moteur s'exerce simultanément sur les roues (transmission) pour le déplacement du tracteur et sur la prise de force pour tracter (traction) l'outil agricole de travail du champ. Ces deux éléments (tracteur et outil agricole) auront un comportement distinct qui est variable en fonction des contraintes exercées. Par exemple, il se peut que, si l'outil agricole est une charrue, les contraintes liées au sol soient plus importantes pour l'outil que pour le tracteur. Et si le choix du tracteur est mal dimensionné par rapport à l'outil agricole, les performances globales seront rapidement dégradées. Dans le passé, la seule manière de répondre à cette question concernait le fait de surdimensionner les performances du tracteur avec, pour conséquence, une augmentation de la consommation de carburant.

Le but de cette étude est de trouver le moyen de déterminer la puissance moteur nécessaire pour fournir les meilleures performances du tracteur en travail agricole avec les outils agricoles adaptés et en fonction des contraintes de sol ou environnementales déterminées. L'utilisation des capteurs est déterminante dans cette étude car elle permet de mesurer les performances du moteur en temps réel lors d'une utilisation agricole du tracteur.

La méthode proposée par cette étude cherche à vérifier s'il est possible

de consulter les valeurs fournies par les capteurs du tracteur en travail agricole pour déterminer exactement quelle était la puissance moteur fournie.

L'étude utilise donc plusieurs capteurs liés au moteur :

- charge moteur ;
- couple instantané ;
- consommation instantanée de carburant ;
- régime moteur.

Un modèle est défini avec les capteurs mentionnés et les valeurs sont mesurées dans des conditions réelles de travail. Ce modèle est ensuite comparé avec des mesures faites sur banc d'essai. Les valeurs mesurées par les capteurs sont donc comparées avec les valeurs mesurées par un dynamomètre. Il apparaît clairement dans cette étude que la manipulation des données fournies par les capteurs, sans application d'une fonction de régression, donne un ensemble de résultats qui sont erronés. La déduction de cette étude est que les valeurs mesurées à faible charge par les capteurs sont celles qui sont les moins précises, en rapport aux valeurs mesurées sur un banc d'essai.

Comment définir la charge d'un tracteur ?

La charge d'un tracteur sera définie par la charge de son moteur. La charge du moteur définit la capacité de celui-ci à produire de l'énergie. Elle est le rapport de la force fournie par le moteur à un certain régime en fonction de la force maximale admise par ce moteur au même régime. Chaque moteur est conçu pour une charge particulière appelée charge nominale ou charge maximale à une vitesse particulière.

2.3 Patinage des pneus

Lors d'un travail de labour, il est important de prendre en compte la nature du sol ainsi que la taille des roues dans le calcul d'estimation de la vitesse du tracteur [14]. Lorsque des modèles d'équations empiriques sont utilisés, il est important de tester ces modèles dans des conditions de terrain multiples. Les résultats de l'étude permettent de définir que les

équations empiriques peuvent être valablement prises en compte lors d'une analyse de valeurs de vitesse ou puissance d'un tracteur en action de travail au champ avec un pourcentage de patinage de roues correctement évalué.

L'étude [15] propose la création d'un capteur permettant de déterminer précisément le pourcentage de patinage des roues motrices d'un tracteur dans différentes configurations d'utilisation. La variation disponible des mesures faites par le capteur en comparaison avec les données calculées est de 2% sur un sol macadamisé. Une équation polynomiale pourrait estimer les tours de roues pour chaque vitesse enclenchée. Par contre, lorsque le tracteur est sur un sol meuble, la variation peut être comprise entre 10% et 44% suivant le travail effectué.

Tant en test sur une route macadamisée (supposée ne pas favoriser le patinage des roues), qu'en test dans un champ, les valeurs mesurées par le capteur proposé peuvent avoir une variation comprise entre 0% et 5% des valeurs réelles.

Cette étude se base sur le principe que l'amélioration des performances d'un tracteur doit passer par un meilleur contrôle des performances de patinage des roues motrices. Pour obtenir les meilleures performances de traction, il est nécessaire que le patinage des roues se situe dans une fourchette de 8% à 15%.

Préalablement à cette étude, plusieurs essais avaient déjà été réalisés pour tenter de mesurer avec précision le patinage des roues, mais sans réel succès. La plupart des techniques utilisées dans ces essais préalables se basaient sur la vitesse théorique mesurée sur un banc d'essai ou sur des méthodes empiriques.

2.4 Couple et régime moteur

Bien que le régime moteur mesuré par le capteur soit proche de la valeur réelle, le couple moteur mesuré, quant à lui, est significativement différent du couple réel [16]. Les actions de travail de champ consommant principalement du carburant sont liées aux opérations de labour, de récolte,

de plantation et de pulvérisation. De fortes variations peuvent être observées entre les différentes marques de tracteurs ou suivant les conditions de fonctionnement. La consommation spécifique de carburant (SFC - Specific Fuel Consumption) a été constatée en augmentation pour les tracteurs avec un niveau de puissance de prise de force plus élevé. La répartition réelle des charges pour l'utilisation d'un tracteur agricole n'est pas connue. Il peut y avoir un pourcentage estimé qui sera compris entre 20% et 30% de l'utilisation du tracteur et qui n'est pas considéré comme étant lié à de l'activité agricole (déplacement vers le lieu de culture, rotation du tracteur en bout de champ). Il est difficile d'évaluer les charges associées aux opérations agricoles. Mais les moteurs peuvent également, pendant les périodes ne résultant pas des opérations agricoles, passer beaucoup de temps à tourner au ralenti. Cette action nécessite, proportionnellement, beaucoup de carburant.

De plus, des charges parasites peuvent augmenter la consommation de carburant d'un tracteur. Il peut s'agir :

- du refroidissement du moteur ;
- de la ventilation de l'habitacle ;
- du frottement du moteur ;
- de la restriction d'admission d'air ;
- de la restriction du système d'échappement ;
- de la pompe à essence ;
- de la pompe à huile ;
- de la pompe à liquide de refroidissement ;
- de l'alternateur ;
- du ventilateur ;
- d'autres accessoires tels que compresseur d'air, climatisation compresseur, etc.

Le changement de vitesse automatique permet également une réduction de la consommation de carburant de 5% à 15% à pleine charge et de 15% à 30% en demi-charge. La possibilité de changement de vitesse automatique permet d'ajouter des paliers de vitesse, réduisant ainsi la plage de tr/min de fonctionnement du moteur, et par la même occasion, la consommation de carburant.

L'estimation du couple moteur d'un tracteur est calculée par le système

OBD (On-Board Diagnostic System) du tracteur, mais aussi via le Bus-CAN. Il est estimé en fonction de la pression et de la durée d'injection du carburant dans le moteur. Néanmoins, cette valeur calculée est surestimée. En effet, le couple moteur calculé au départ du Bus-CAN est surévalué en raison de la résistance interne du moteur et induit une estimation des émissions dont la résultante est trop faible. Il faut donc pouvoir estimer correctement cette résistance pour la prendre en compte dans le calcul du couple moteur [17].

Dans cette étude, la mesure de consommation de carburant est effectuée sur banc d'essai via les données de diagnostic. La consommation de carburant est déterminée en rapportant les valeurs enregistrées du régime moteur. La confirmation des résultats est faite en vérifiant les valeurs d'émissions relevées simultanément. L'exactitude des valeurs est donc confirmée. Ces mesures ont été réalisées à plusieurs régimes moteur distincts. Les pertes de combustion dans un moteur dépendent principalement du régime moteur et, dans une moindre mesure, de la charge du moteur. Il est supposé que les pertes internes du moteur sont constantes pour un régime moteur donné. L'écart entre les régimes moteurs définis et mesurés peut varier de 3,4% à 6,8%.

Cette charge moteur est disponible dans la consultation des données fournies par le Bus-CAN. Une analyse de la valeur charge fournie par le Bus-CAN du tracteur a été réalisée [18], le but étant de vérifier la précision du couple moteur net telle que fournie par les messages du Bus-CAN. Cette analyse permet de mettre en évidence que le pourcentage de charge, en charge réelle, n'est pas équivalent à la charge relevée par le Bus-CAN en comparaison avec un dynamomètre lors d'un test en banc d'essai. Cette étude souligne qu'il n'est pas possible de savoir si les capteurs sont étalonnés et si les résultats des valeurs calculées ne peuvent être vérifiées. Aussi, les mesures empiriques prenant en compte toutes les plages de valeurs devraient également pouvoir être proposées pour tous les types d'action du tracteur.

Une régression linéaire entre la mesure du nombre de tours minutes du moteur via un dynamomètre et la valeur fournie par le Bus-CAN prouve que ces deux valeurs ont une corrélation parfaite.

Cette étude fournit une preuve complémentaire de l'imprécision de la valeur du couple moteur donnée par le Bus-CAN. Le couple calculé est en

général plus élevé que le couple réellement mesuré par le dynamomètre. Malgré cela, la régression entre les valeurs mesurées et celles fournies par le Bus-CAN permet de déterminer une forte corrélation entre les deux valeurs mais avec un décalage pour lequel le couple fourni par le Bus-CAN est toujours plus élevé que le couple mesuré par le dynamomètre. Ce phénomène peut être dû aux calculs établis par le constructeur mais également par la faible résolution (nombre de fois par seconde) dans les mesures fournies par le Bus-CAN. De plus, cette étude suggère également de prévoir des tests similaires pour des moteurs de marques et de modèles différents.

2.5 ECU et Etalonnage

Cette fiabilité est aussi étudiée dans l'étude [19]. Il est supposé que les méthodes d'acquisition des données peuvent induire des erreurs dans les valeurs fournies. Cette analyse a été déjà proposée plus haut (Samuel E. Marx, 2015) [9]. Néanmoins, la question posée cette fois est de savoir s'il est possible de réduire le volume des données.

Les données de trame du Bus-CAN peuvent être problématiques de par le grand volume de données. Plusieurs formats de données ont, à nouveau, été analysés dans cette étude pour comparer leur taille globale. Il s'agit d'une trame dont les valeurs résultantes sont des moyennes de valeurs fournies ou bien des jeux de données de forme d'ondes. Les valeurs mesurées par la méthode des trames et par la méthode de forme d'ondes a permis de détecter que les deux techniques offrent des valeurs proches mais avec un moins grand volume en ce qui concerne les données de forme d'ondes.

La technique d'enregistrement des trames du Bus-CAN reste néanmoins la technique la plus simple à mettre en oeuvre et la plus utilisée. L'enregistrement en forme d'ondes est une technique de rééchantillonnage des données de trames Bus-CAN en forme d'ondes à une fréquence d'échantillonnage fixe (NI-TDMS [20]) via le logiciel LabVIEW. Ces données sont moyennées à une fréquence d'échantillonnage de 2hz.

Cinq jeux de données sont choisis dont trois jeux de données ayant des fréquences distinctes : vitesse du moteur, taux de carburant et température

de carburant. Ces trois types de données sont ramenés à une fréquence de 1hz pour une première comparaison. Ensuite ces données sont comparées avec des valeurs fournies par un test dynamométrique, mais également rééchantillonnées à différentes fréquences pour une deuxième comparaison. Seule la vitesse du moteur a montré des valeurs différentes entre les différents types de relevés. Néanmoins une bonne synchronisation des fréquences d'échantillonnage permet de réduire les erreurs entre les différentes méthodes de relevés tout en constatant que la méthode de relevé par forme d'ondes réduit de 6% la taille des fichiers en rapport à ceux produits par les trames J1939. Il n'est donc pas possible d'en conclure que la méthode utilisée pour relever les données pourrait introduire des erreurs significatives.

Comme mentionné plus haut, plusieurs valeurs fournies dans le data logger connecté au Bus-CAN sont des valeurs calculées dans ECU au départ des informations fournies par les capteurs. Il est intéressant également de prendre en compte la réalisation d'un ECU [21] pour vérifier s'il peut introduire des erreurs lors de la capture des données.

Cette étude se base sur AUTOSAR (AUTomotive Open System ARchitecture) qui est une architecture logicielle standardisée et ouverte pour les unités de contrôle électroniques ECU. AUTOSAR fournit un ensemble de spécifications. Celles-ci définissent les modules logiciels de base, les interfaces applicatives et elles élaborent une méthodologie de développement commune reposant sur un format d'échange standardisé.

A travers cette étude, il est constaté que le débit du Bus-CAN est à 100% lorsqu'il atteint 800kbps (Kbits par seconde) et que les pertes d'informations commencent à 60% de la charge maximale du débit pour atteindre un maximum de 14% de pertes lorsque le débit est maximal. Ce qui peut représenter un élément non négligeable de source d'erreurs lors de la capture des données du Bus-CAN.

Néanmoins, un ECU peut être étalonné [22]. La méthodologie CCP (CAN Calibration Protocol) est une norme de calibration qui est gérée par le groupe de travail Standardization of Application/Calibration Systems (ASAP - Arbeitskreis zur Standardisierung von Applikationssystemen). Ce groupe de travail propose une stratégie basée sur le réseau Bus-CAN pour l'étalonnage des calculateurs ECU. Cette proposition vient compléter les

deux premières propositions disponibles via le Bus-CAN que sont la mesure et le diagnostic.

Comme constaté à travers la lecture de ces études, plusieurs sources d'erreurs sont possibles lors de la captures des données fournies par le Bus-CAN.

Il peut s'agir d'erreurs venant :

- du capteur en lui-même ;
- de l'ECU qui calcule certaines valeurs sur base des données qu'il reçoit de plusieurs capteurs ;
- de la synchronisation des éléments composant les modules de capture des données ;
- de la charge du réseau Bus-CAN et des éléments qui composent le réseau ;
- du volume des données disponibles lors d'un travail agricole d'un tracteur ;
- ...

Chapitre 3

Analyse de données disponibles

Sur base des constats effectués à la lecture des études préalablement faites sur le sujet, une vérification de la fiabilité des données en utilisant des jeux de données à notre disposition est proposée.

Deux fichiers distincts contenant les relevés de tracteurs sont utilisés pour cette analyse.

Le premier relevé de données utilisé est fourni par un tracteur se déplaçant sur la route. Ce relevé sur route permet de ne pas prendre en compte les contraintes liées à la nature du sol qui pourraient introduire des erreurs difficiles à évaluer correctement (notamment dues à une zone de terrain plus souple introduisant une augmentation du patinage des roues).

Le deuxième fichier est relevé sur un tracteur ayant une activité de travail au champ. Ce second fichier permettra de comparer le comportement des valeurs avec le premier relevé suivant le type d'action du tracteur.

Dans un tracteur agricole, il est possible de différencier les capteurs selon deux types :

- capteurs fournissant des valeurs à état (vrai - faux, ouvert – fermé, etc) ;
- capteurs fournissant des valeurs numériques.

Dans cette étude, une attention particulière est mise sur quelques capteurs, fournissant des valeurs numériques, qui peuvent être sujets à la

production d'erreurs.

Les données les plus susceptibles de contenir des erreurs de précisions sont liées :

- soit au déplacement du tracteur ;
- soit à la traction fournie par le tracteur.

La condition posée est que les données prélevées pour cette vérification soient fournies par un tracteur en condition réelle d'utilisation, car :

- ces données sont facilement disponibles en grandes quantités ;
- pour permettre l'utilisation des données pour déterminer de nouveaux modèles mathématiques, il est nécessaire que les vérifications de fiabilité soient faites avec des valeurs collectées dans des conditions réelles d'utilisation.

Le choix retenu est le capteur de régime moteur (nombre de tours/minutes du moteur). En effet, les données du régime moteur peuvent être utilisées dans un grand nombre de cas de figure : pour calculer la consommation, la puissance ou performance du tracteur, etc [8 - 9 - 11 - 13 - 16 - 17 - 18].

Une étude [23] précise que, sur route, les données du régime moteur relevées par Bus-CAN sont fortement corrélées avec les données mesurées sur un banc d'essai (il ne s'agit pas de données fournies par un tracteur mais par un camion).

Pour comparer les valeurs fournies par le régime moteur, il est naïvement supposé que la vitesse du tracteur peut servir de référence. Ainsi, lorsque le tracteur se déplace à vitesse constante, il est supposé que cette vitesse constante, associée à l'enclenchement d'un numéro de vitesse défini dans la boîte de vitesse, devra correspondre à un certain régime moteur. Il s'agira, pour le déplacement du tracteur, de comparer la vitesse avec le nombre de tours/minutes du moteur.

Il est nécessaire de définir des capteurs de référence pour analyser le capteur de régime moteur.

Etant donné qu'une vérification avec des valeurs d'utilisation réelle d'un tracteur est effectuée, il est intéressant de se demander si d'autres valeurs de

capteurs disponibles dans le jeu de données sont utilisables comme valeurs de référence.

Les capteurs de référence choisis, pour le calcul de la vitesse, sont les capteurs de positionnement GPS. Ce choix est motivé par le fait qu'il existe deux types distincts de valeurs GPS fournies dans le jeu de données, mais également une vitesse calculée sur base des positions GPS.

Les valeurs de position GPS, relevées dans le fichier disponible, seront utilisées. Elles permettront de déterminer la vitesse du tracteur dans un intervalle de temps déterminé. Pour cela, il sera nécessaire de convertir les positions relevées chaque seconde en valeur de déplacement.

Néanmoins, quelle est la fiabilité des valeurs fournies par un GPS ?

3.1 Fiabilité de positionnement GPS

Il est nécessaire de savoir si l'utilisation de la position GPS est une donnée fiable à exploiter. Plusieurs études mentionnent les types d'erreurs et les méthodes de correction utilisées.

Ces erreurs peuvent être dues :

- à la couche de troposphère (l'humidité et la variation de pression modifient le trajet du signal satellite) ;
- à la couche de l'ionosphère (ensemble de couches ionisées qui modifient la vitesse de propagation du signal satellite) ;
- au déplacement du satellite sur son orbite ;
- aux réflexions des bâtiments et autres surfaces au sol ;
- ...

Une étude [24] porte sur le fait que les données de distance calculées par le GPS ne sont pas suffisamment fiables (précision à 100m) à elles seules pour être utilisées en agriculture de précision. Il est nécessaire d'y ajouter des corrections de distances capturées par le biais d'une autre source pour augmenter la fiabilité de positionnement (solution DGPS - Differential Global Positioning System). Cette correction est fournie par une antenne fixe positionnée à un endroit dont les coordonnées sont connues. Elle permet

une très grande précision mais requiert une grande attention pour sa mise en oeuvre et son utilisation (utilisation de GPS spécifiques). Les précisions sont de l'ordre du centimètre mais la technique n'est pas très robuste. Malgré cela, Il est important de souligner que des erreurs peuvent aussi se produire via un DGPS, notamment si l'alignement de l'antenne n'est pas réalisé précisément.

L'erreur maximale prévue, pour une solution DGPS à code C/A [25] pourrait être estimée à 1 m horizontalement. Même si un GPS haute précision à code C/A capable de recevoir des corrections bénéficie d'une bonne précision, il reste un processus complexe à mettre en oeuvre. Les positions mesurées par les solutions DGPS prennent plus de temps à être mises en oeuvre.

La mesure de distance GPS par le biais d'un relevé de la position dans un intervalle de temps régulier ne permet pas de considérer que le calcul de la vitesse soit précise car il peut exister une différence de vitesse allant jusqu'à 0,25m/s (0,9 km/h). Le GPS à code C/A, complété d'une correction de positionnement, ne permet donc pas de mesurer des positions instantanées correctes, elles-mêmes ne permettant pas d'estimer des distances et des vitesses de fonctionnement précises. Une amélioration proposée peut être obtenue si l'on effectue une moyenne de plusieurs relevés distincts pour chaque point mesuré.

L'étude [26] permet de constater que, sur terrain accidenté, les mouvements de l'antenne, placée sur le tracteur, et de son centre de gravité peuvent introduire des erreurs de positionnement du tracteur. Ces erreurs affectent la vitesse instantanée mais, lors de long déplacement, ces erreurs peuvent être corrigées par le biais de moyennes mobiles ou de corrections mathématiques. Néanmoins, cette vitesse fournie par le GPS peut être utilisée pour détecter les changements de vitesse au sol et la puissance développée. La force de résistance totale peut être calculée à partir des mesures du Bus-CAN et de la vitesse obtenue par le GPS. En outre, ces méthodes de relevés des données sont simples à mettre en oeuvre et peu coûteuses.

Il existe également une erreur liée aux mauvaises estimations de calcul de distances notamment parce que les estimations de distances dans les courbes peuvent être erronées [27]. Lors d'une estimation de distance importante

(plusieurs kilomètres), le déplacement peut ne pas être égal à la distance. Le choix de routes sinueuses rendra le déplacement plus long que la distance à parcourir.

Contrairement à cette étude, étant donné qu'uniquement de courtes distances (équivalentes à quelques mètres en raison de la fréquence des relevés) sont utilisées, chaque déplacement est supposé équivalent à la distance puisque :

- le déplacement est en ligne droite ;
- la distance mesurée se fait entre deux positions relevées chaque seconde.

Il est néanmoins nécessaire de tenir compte des opérations de retournement du tracteur en bord de champ. Dans cette d'analyse, ce point ne sera pas à prendre en compte dans le premier relevé de mesures puisque le déplacement est en ligne droite. Néanmoins, il devra être pris en compte lors de l'analyse du deuxième fichier.

Deux types de correction [28] des erreurs énumérées plus haut sont disponibles :

- Le SBAS (Satellite base Augmentation System). Le satellite reçoit les données de correction et les transmet au récepteur GPS en même temps que le signal de base ;
- Le LBAS (Local Based Augmentation System). Utilise une transmission cellulaire pour envoyer les corrections au récepteur GPS.

L'étude [29] permet de constater qu'il peut être possible de mesurer la vitesse de déplacement d'une personne par le biais d'un GPS non différentiel. Les vitesses fournies par le GPS lors d'un parcours rectiligne étaient étroitement corrélées avec les vitesses mesurées. L'erreur moyenne de position était de 34cm par rapport au point réel. Etant donné que pour des opérations en travail agricole, les vitesses ne sont pas plus élevées que le déplacement d'un homme à pied, il peut rester intéressant de vérifier si ces relevés donnent de bons résultats.

Comme vu plus haut, les GPS différentiels (DGPS) sont les GPS qui bénéficient d'une correction d'erreur par le biais d'une antenne satellite fixe positionnée à un endroit prédéfini et pour laquelle la position exacte

est connue. Cette antenne est couplée à un dispositif d'émission VHF permettant d'envoyer des signaux de corrections aux GPS mobiles. Le GPS différentiel mobile doit donc être équipé d'une méthode de réception complémentaire VHF permettant de recevoir les signaux de correction. Les GPS non différentiels, quant à eux, ne bénéficient pas d'une grande précision de correction mais ils sont moins complexes à mettre en oeuvre car ils ne nécessitent pas d'antenne fixe permettant d'envoyer les signaux de correction.

Le GPS non différentiel peut offrir une précision de calcul de vitesse de déplacement uniquement à vitesse constante le long de trajets en ligne droite. Une forte corrélation est déterminée entre les vitesses calculées de déplacement d'un humain sur une trajectoire rectiligne et la vitesse mesurée par le GPS. Il est donc raisonnable d'estimer qu'il est possible d'obtenir les mêmes résultats pour un tracteur en travail agricole.

Enfin, une étude [30] tente de répondre à la question de l'estimation de la vitesse de déplacement d'un véhicule par le biais d'une application GPS installée sur le smartphone d'un occupant du véhicule. La raison première de cet essai est de pouvoir bénéficier d'une technologie (GPS embarqué dans un smartphone) simple et peu coûteuse avec une forte pénétration du marché.

Dans cette étude, la vitesse fournie par le GPS est comparée avec le système de diagnostic OBD-2 embarqué dans le véhicule. Celui-ci fournit également une estimation de la vitesse. Le principe de cette étude est de déterminer les changements brusques de vitesse pendant le déplacement. La principale contrainte dans cette étude est liée au fait que le smartphone doit être fixe et toujours situé au même endroit dans le véhicule lorsque celui-ci est en mouvement. Bien évidemment, il est supposé que la précision des valeurs fournies par un smartphone ne peut être fiable. Pour effectuer ces corrections de précisions, deux filtres sont introduits dans l'application développée pour le smartphone. Il s'agit de filtres de moyennes mobiles avec des coefficients de pondérations normalisés.

De plus, à titre de comparaison, les données OBD-2 du véhicule sont distribuées à l'application du smartphone via un adaptateur Bluetooth. De manière simultanée, les GPS enregistrent leurs positions et capturent la vitesse instantanée. La vitesse instantanée mesurée par le GPS, lorsqu'elle est constante, est inférieure d'environ 2 à 4 km/h à la vitesse fournie par les

éléments de diagnostic OBD-2.

3.2 Hypothèses de départ

Pour débiter l'analyse des relevés, quelques hypothèses de départ sont fixées. Elles vont permettre de définir ce qui est attendu comme comportement et comme résultats.

Existe-t-il une forte relation entre la vitesse du tracteur et le nombre de tours/minutes du moteur ?

Se peut-il que lorsque la vitesse augmente, le nombre de tours/minutes du moteur augmente également ?

Se peut-il que, pour une vitesse constante de déplacement, il y ait des variations du nombre de tours/minutes ?

Hypothèses :

- 1. il est tout d'abord supposé qu'il existe une relation directe entre le régime du moteur exprimé en tours-minutes et la vitesse du tracteur ;
- 2. la vitesse du tracteur peut être calculée par le biais d'un ensemble de positions GPS récoltées chaque seconde lors de l'utilisation du tracteur. Une opération de calcul de distances sera réalisée au départ des relevés de positions ;
- 3. il est également supposé que le relevé de la position GPS peut être soumise à erreurs de par le principe de fonctionnement des satellites émetteurs et des récepteurs GPS. Néanmoins, si la vitesse est constante durant le temps du déplacement, il sera estimé que les corrections de positionnement GPS sont fiables ;
- 4. puisque le premier jeu de données du tracteur se fait sur route, il est supposé que le patinage des roues est à 0.
- 5. il est enfin supposé que, puisque le tracteur est sur route ouverte, il n'existe pas de force appliquée au moteur en relation avec la traction d'un outil agricole. A l'inverse, il sera intéressant de constater quelles sont les valeurs de positionnement GPS et de régime moteur pour le deuxième fichier pour lequel le tracteur sera en phase de travail.

3.3 Choix des données pour cette première analyse

Les capteurs intéressants dans la première phase d'analyse sont :

- le régime moteur (en nombre de tours/minutes) ;
- les positions GPS pour le calcul des distances.

Le nombre de tours-minutes du moteur

Le PGN correspondant au nombre de tours/minutes est : 61444 et le SPN est 190. Son libellé exact est "Engine Speed". Il correspond à la colonne "Isobus : : EEC1_TECU : : EngSpeed" du premier fichier et à la colonne "EngSpeed" dans le deuxième fichier.

Les positions GPS

Les positions GPS relevées dans chaque fichier proviennent de deux capteurs distincts.

Les deux capteurs sont distingués en les nommant respectivement GPS "GPS-1" pour le premier, et "GPS-2" pour le deuxième.

GPS-1 :

Pour le premier relevé, le GPS-1 correspond aux colonnes libellées :

- j1939 : :VP : :Longitude ;
- j1939 : :VP : :Latitude.

Pour le deuxième fichier, le GPS-1 correspond aux colonnes libellées :

- Longitude ;
- Latitude.

GPS-2 :

Pour le premier relevé, le GPS 2 correspond aux colonnes libellées :

- GPS : :GPS1.Longitude ;
- GPS : :GPS1.Latitude.

Pour le deuxième relevé, il correspond aux colonnes libellées :

- GPSLongitude;
- GPSLatitude.

Enfin, il existe également une vitesse "GPSspeed" qui est calculée par les positions GPS. Dans le premier fichier, le libellé de la colonne est "GPS : :GPS1.Speed" et dans le deuxième fichier, la colonne correspondante est nommée "GPSSpeedOverGround".

Le détail des appareils GPS utilisés n'est pas fourni mais, pour un des deux capteurs, le PGN correspondant est : 65267 et son SPN est

- Latitude : 584;
- Longitude : 585.

3.4 Choix des fichiers de données

Caractéristiques du premier relevé d'un tracteur

Pour le premier relevé, le tracteur utilisé est :

- Tracteur : Massey Fergusson 7619 de 2013;
- Puissance : 185hp de puissance nominal, 220hp de puissance max, transmission power shift.

Trajet effectué lors du premier relevé ;

- position de départ du tracteur : Chaussée de Namur 142, 5030 Gembloux, Belgique - Latitude :50.551237 - Longitude :4.708831.
- position d'arrivée du tracteur : Carrefour dit « Didi » Chaussée de Namur sur la Nationale 4 à Bovesse, Belgique - Latitude : 50.522145 - Longitude : 4.759286.

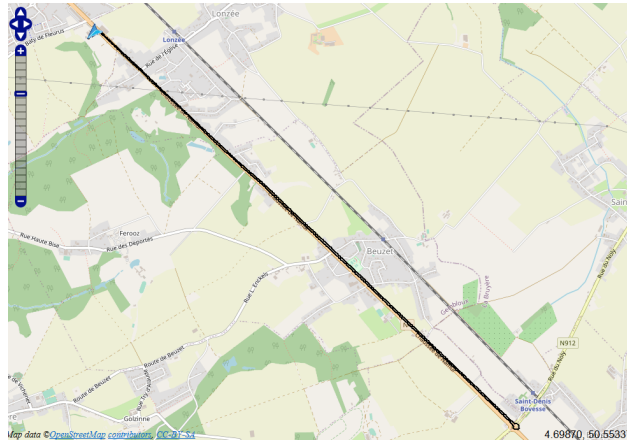


FIGURE 3.1 – **Trajet tracteur** – Trajet du tracteur du premier relevé.

La distance estimée est de $\pm 4,3$ Km pour 4 min de trajet à 90 Km, ce qui signifie 7 min à 43 km. Le fichier final dont nous disposons est de 620 lignes mais seulement les lignes allant de la 137 à la ligne 557 contiennent de valeurs de déplacement, ce qui donne 420 lignes. 420 lignes pour 420 secondes divisées par 60 = 7 minutes de trajet.

Caractéristiques du deuxième relevé d'un tracteur

Le deuxième fichier, quant à lui, contient des valeurs de déplacement sur un champ. Le détail du véhicule utilisé n'est également pas fourni. Le champ est sur terrain plat mais le tracteur fait plusieurs demi-tours lors de l'enregistrement de son trajet.

La position de départ se situe à la Latitude : 50.994213 et la Longitude : 3.786952 alors que la position de fin est à la Latitude : 50.993629 et la Longitude : 3.785978.

Le positionnement étant dans un champ, il n'existe pas d'adresse définie.

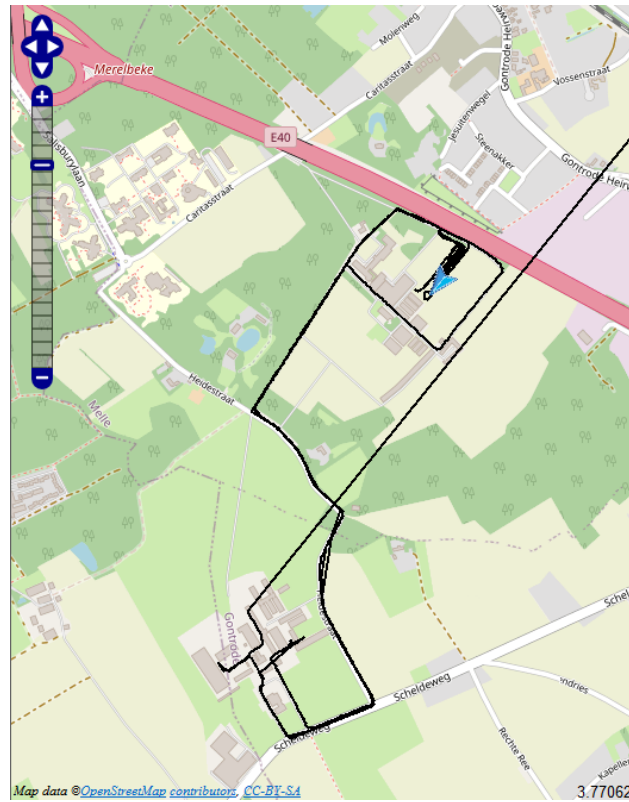


FIGURE 3.2 – **Trajet tracteur** – Trajet du tracteur du deuxième relevé.

Le relevé des données ne concerne que la portion de déplacement dans le champ.

Chapitre 4

Méthodologie de l'analyse des données

4.1 Collecte des données.

Les données utilisées sont déjà disponibles au format CSV. Pour ce faire, les fichiers collectés au départ du tracteur ont déjà subi un premier traitement qui leur permet de convertir les données brutes, du format original en hexadécimal, vers un format lisible. Le détail de cette opération n'est pas décrit car il est purement technique, mais un logiciel comme « Vector » peut effectuer ce genre d'opération.

Comme premier fichier d'analyse nous utilisons un fichier contenant +/- 144000 lignes de données. Le fichier est également constitué de 69 colonnes. Il s'agit d'un fichier contenant approximativement 240 enregistrements par seconde pour un total d'un peu plus de 600 secondes.

Comme deuxième fichier utilisé pour l'analyse, nous utilisons un fichier contenant 3302 lignes pour 66 colonnes. Ce deuxième fichier contient approximativement 5 enregistrements par seconde, pour un total de 662 secondes.

Chaque colonne, exceptée la première, des deux fichiers représente un capteur et les lignes représentent la valeur mesurée pour chacun de ces capteurs à un instant T défini dans la première colonne (Time[s]).

4.1.1 Nettoyage du fichier contenant les données

Le premier travail consiste donc à ne garder qu'une ligne par seconde.

La valeur de temps mentionnée dans la colonne [Time(s)] étant au format décimal, seule la première valeur la plus proche de x,004 (millième de seconde) sera prise en compte pour le premier fichier. Celui-ci n'a pas été traité préalablement pour établir une valeur médiane pour des intervalles de temps réguliers. Au total, le fichier obtenu a une taille de 652 lignes.

Le deuxième fichier contient 3200 lignes mais avec un échantillonnage à 5 Hz (5 lignes par seconde). Contrairement au premier fichier, il y a dans ce deuxième fichier une pondération des valeurs pour correspondre au timing exact de décimale de temps.

Chaque décimale de temps est à 0,0 seconde, puis à 0,2 seconde, puis à 0,4 seconde, puis à 0,6 seconde, et enfin à 0,8 seconde. Mais il n'existe pas l'ensemble de ces 5 lignes pour chaque seconde. Pour certaines secondes, il n'existe que 4 lignes. Il est donc essentiel de ne conserver qu'une seule ligne par seconde. Celle qui sera gardée est celle qui est à l'intervalle 0,0 seconde.

Les colonnes retenues dans le premier fichier sont :

- 1ère colonne : Time[s] qui est le temps écoulé entre chaque relevé. Il est exprimé en secondes avec décimales.
- 2ème colonne : GPS : :GPS1.Latitude[degrés] qui est la position de la latitude GPS exprimée en degrés décimaux ;
- 3ème colonne : GPS : :GPS1.Longitude[degrés] qui est la position de la longitude GPS exprimée en degrés décimaux ;
- 5ème colonne : GPS : :GPS1.Speed[m/s] qui est la vitesse calculée du tracteur exprimée en mètres par seconde ;
- 6ème colonne : Isobus : :EEC1_TECU : :EngSpeed[rpm] qui est le nombre de tours/minutes du moteur exprimé en rpm ;
- 63ème colonne : J1939 : :VP :Longitude[deg] qui est la position de la longitude J1939 exprimée en degrés décimaux ;
- 64ème colonne : J1939 : :VP :Latitude[deg] qui est la position de la latitude J1939 exprimée en degrés décimaux.

Les colonnes retenues dans le deuxième fichier sont :

- 1ère colonne : Time[s] ; qui est le temps écoulé entre chaque relevé. Il est exprimé en secondes avec décimales.
- 5ème colonne : GPSTime[decimal] ; qui est la position de la longitude GPS exprimée en degrés décimaux ;
- 6ème colonne : GPSTime[decimal] qui est la position de la latitude GPS exprimée en degrés décimaux ;
- 10ème colonne : GPSSpeedOverGround ; qui est la vitesse calculée du tracteur exprimée en mètres par seconde ;
- 12ème colonne : EngSpeed[rpm] ; qui est la vitesse du moteur exprimée en tours/minutes ;
- 64ème colonne : Longitude[deg] qui est la position de la longitude GPS exprimée en degrés décimaux ;
- 65ème colonne : Latitude[deg] qui est la position de la latitude GPS exprimée en degrés décimaux ;

4.2 Calcul de la vitesse

La méthode choisie pour obtenir la vitesse du tracteur se base sur l'utilisation des positions GPS relevées chaque seconde. Une distance de déplacement (en mètres) pour chaque seconde écoulée est ainsi obtenue.

Le calcul de la vitesse de déplacement via l'utilisation du GPS a été réalisé à de nombreuses reprises, notamment via l'étude d'Andrew D. Townshend et al. [29] concernant le déplacement d'un être humain.

Bien que plusieurs études mentionnent le fait que la précision du GPS ne soit pas fiable, il sera supposé que, pour chaque seconde, à vitesse constante estimée par le capteur GpsSpeed, si la vitesse calculée au départ des positions GPS reste constante, l'erreur de positionnement sera faible.

N'est pris en compte que la mesure de position GPS qui peut être suffisamment précise lorsque la vitesse de déplacement reste constante. Une fois que la vitesse de déplacement varie, la précision peut rapidement diminuer.

Il peut être prévu, pour obtenir des résultats dans toutes les plages

de fonctionnement, d'utiliser des équations empiriques permettant de couvrir l'ensemble des valeurs voulues. Néanmoins, ce travail est fastidieux et doit être appliqué pour chaque type de véhicule (chaque marque, modèle, etc). Malgré cela, comme constaté préalablement, au départ des analyses d'études résumées plus haut, les valeurs constantes durant une plage de temps déterminée sont plus susceptibles de donner de bons résultats.

La méthode utilisée pour la comparaison des données se base sur le schéma suivant :

- collecte des données générées par un tracteur lors de son utilisation avec premier traitement des données pour les rendre lisibles dans un tableau de type CSV (opération déjà effectuée préalablement) ;
- établissement d'un choix de capteurs pour en déterminer la vitesse du tracteur ;
- réduction du nombre de lignes à 1 ligne/seconde ;
- calcul des valeurs de déplacement et valorisation de la vitesse moyenne comme mesure de référence.

Un script "R" sera utilisé pour calculer la vitesse au départ des positions GPS.

4.2.1 Script R de calcul de la vitesse

Une fois que l'ensemble des lignes contenant les valeurs par seconde est obtenu, la vitesse existante entre chaque relevé de position GPS est déterminée.

Pour ce faire, un script "R" est utilisé. Il permet de calculer rapidement cette vitesse. Tout d'abord, les valeurs en entrée sont fournies au script par le biais d'une lecture dans le fichier CSV. Les colonnes à lire dans le fichier sont déterminées. Le nom de ces colonnes est mentionné dans les variables locales "lon" et "lat". La structure du fichier CSV est ensuite mentionnée pour que le script puisse le parcourir. Une boucle sur le nombre de lignes nécessaires est créée et le calcul de chaque distance du point courant avec le point suivant est établi.

Deux méthodes sont utilisées pour le calcul de la distance. Ces deux méthodes sont disponibles dans le langage "R". Il s'agit de la méthode Harversine [30] et de la méthode distGeo. Les résultats des deux méthodes

sont comparés pour en vérifier la similitude. De courtes distances sont manipulées entre les points, il est donc essentiel de s'assurer de la meilleure formule de calcul de distances.

Enfin, les résultats sont exportés dans un nouveau fichier.

Le script "R" est représenté ci-dessous.

```
library(geosphere)
library(dplyr)
library(disk.frame)

# Répertoire courant pour chercher le fichier setwd("E :/Master 60
informatique UNamur/Memoire/Fichiers calculs/Fichiers de base/2eme")

# Nombre de lignes du fichier
NbrLines <- 662

# -----
# CALCUL SUR BASE DE GPS-2
# -----

lon = "Longitude.deg."
lat = "Latitude.deg."

#Lecture dans le fichier XLS pour récupérer les coordonnées latitude-
longitude
fileName = c("Releve tracteur par seconde.csv")

tractorPosition <- read.csv(file = "Releve tracteur par seconde.csv",
header = TRUE,
sep = ";",
skipNul = TRUE,
```

```

stringsAsFactors = FALSE),
head(tractorPosition)

for (x in 1 : NbrLines) {
line1 = tractorPosition[x, ]
line2 = tractorPosition[x+1, ]

lon1 = (tractorPosition[c(x),lon ])
lon1num = as.numeric(gsub(",",".", lon1))
lat1 = (tractorPosition[c(x),lat ])
lat1num = as.numeric(gsub(",",".", lat1))
lon2 = (tractorPosition[c(x+1), lon ])
lon2num = as.numeric(gsub(",",".", lon2))
lat2 = (tractorPosition[c(x+1), lat ])
lat2num = as.numeric(gsub(",",".", lat2))

# Calcul distance via méthode Haversine
dist2[x] <- distm(c(lon1num[1], lat1num[1]), c(lon2num[1],lat2num[1]),
fun=distHaversine)
print(dist2[x])

# Calcul distance via méthode distGeo
p1 <- c(lon1num[1], lat1num[1])
p2 <- c(lon2num[1], lat2num[1])

print(p1)
print(p2)
dist3[x] <- distGeo(p1,p2)
print(dist3[x])
}

write.table(dist2, file="exportDistance_GPS-2.csv", sep=";", dec=',')

write.table(dist3, file="exportDistance_GPS-2-MGEO.csv", sep=";",
dec=',')

```

Un fois les données de distances obtenues suite au calcul des positions GPS, elles peuvent être comparées avec les mesures de vitesses obtenues dans le set de données. Il est à noter que les valeurs de vitesse instantanée disponibles dans le fichier sont le fruit d'un calcul effectué par un des deux modules GPS. La première consultation des données sera visuelle et portera sur la comparaison des vitesses calculées par le script "R" avec la vitesse calculée par un module GPS et dont les résultats sont dans le fichier.

Sont tout d'abord comparés, pour chaque relevé (tracteur sur route et tracteur au champ) et chaque capteur GPS (GPS-1 et GPS-2), les résultats de calcul par la méthode "Harvesine", et ensuite via la méthode "DistGeo".

Ensuite, le capteur GPS dont la fiabilité est la plus grande sera sélectionné.

4.2.2 Vérification des vitesses calculées pour le premier relevé

Les graphiques suivants présentent les relevés des différents GPS pour le premier relevé. Il s'agit de graphiques présentant la vitesse du tracteur en fonction de la durée du déplacement.

Pour les 4 figures suivantes, la comparaison est faite, dans la durée, entre les valeurs calculées sur base des positions GPS via le script "R" (courbes en bleu) et les valeurs calculées qui sont déjà fournies dans le fichier CSV (courbes en orange).

Relevé sur route pour GPS-1

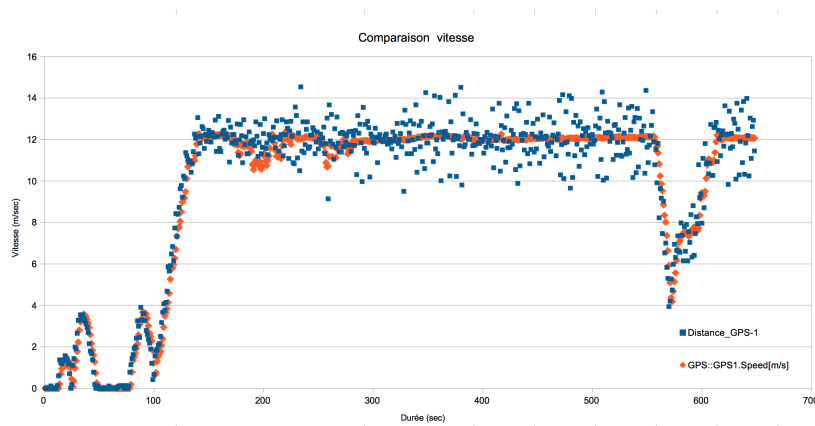


FIGURE 4.1 – **Vitesse calculée suivant la durée du trajet** – Vitesse du tracteur dans la durée pour le 1er fichier et le capteur GPS-1, calculée par la méthode Haversine.

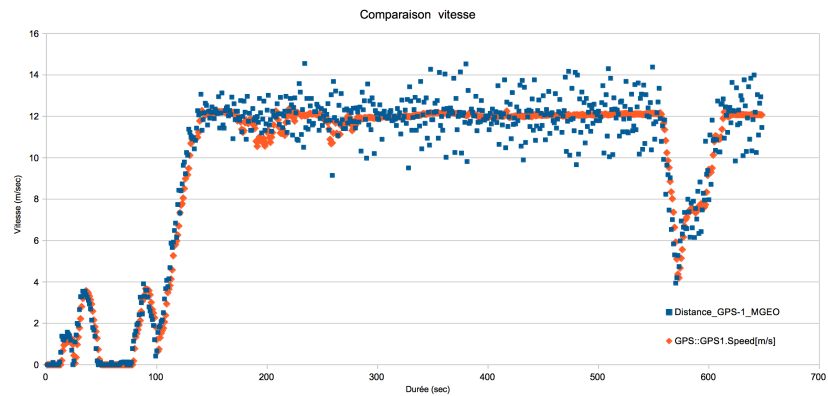


FIGURE 4.2 – **Vitesse calculée suivant la durée du trajet** – Vitesse du tracteur dans la durée pour le 1er fichier et le capteur GPS-1, calculée par la méthode DistGeo.

Relevé sur route pour GPS-2

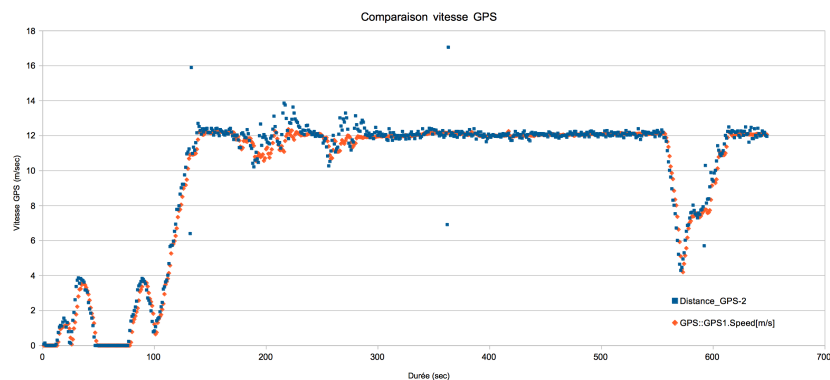


FIGURE 4.3 – **Vitesse calculée suivant la durée du trajet** – Vitesse du tracteur dans la durée pour le 1er fichier et le capteur GPS-2, calculée par la méthode Haversine.

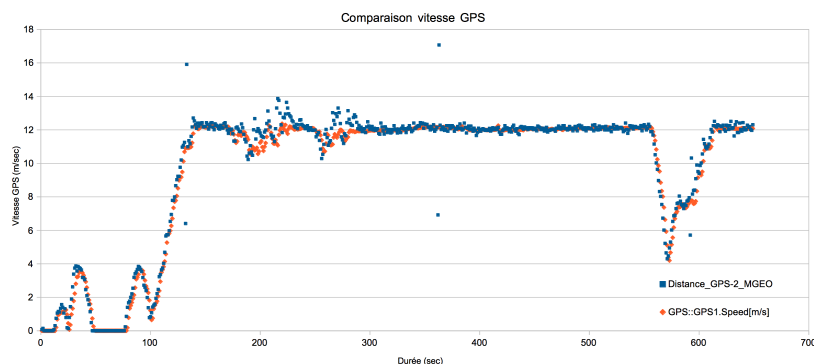


FIGURE 4.4 – **Vitesse calculée suivant la durée du trajet** – Vitesse du tracteur dans la durée pour le 1er fichier et le capteur GPS-2, calculée par la méthode DistGeo.

Constatations

Plusieurs points sont constatés grâce à la représentation graphique des jeux de données obtenus :

- tout d’abord, la courbe de vitesse n’est pas constante dans le temps, ce qui est normal de par le fait que le tracteur doit avoir une accélération

- avant d'atteindre sa vitesse maximale ;
- ensuite, il est constaté que la courbe de GPS-2 fournit des points moins dispersés que la courbe de GPS-1 ;
- enfin, pour chacun des deux relevés GPS, la différence entre la méthode de calcul "Harversine" et la méthode de calcul "DistGeo" ne fournit pas de différences notables.

Les valeurs fournies par le GPS-1 et par le GPS-2 sont ensuite analysées. Afin de faciliter les analyses ultérieures, il sera intéressant de ne tenir compte que d'un seul des deux capteurs GPS.

La comparaison est faite entre les valeurs disponibles dans les deux GPS pour tenter de déterminer quel capteur est le plus fiable.

Pour ce faire, la valeur de vitesse fournie dans le fichier de base est tout d'abord comparée avec la valeur calculée par la méthode "Harversine" du script "R" afin d'évaluer la précision.

Le choix de la régression se porte sur une régression linéaire puisqu'il est supposé qu'il existe une relation directe entre les deux mesures. Il est supposé que la valeur de vitesse augmente de la même manière pour les données disponibles dans le fichier et pour les valeurs calculées. La variable explicative est la valeur disponible dans le fichier.

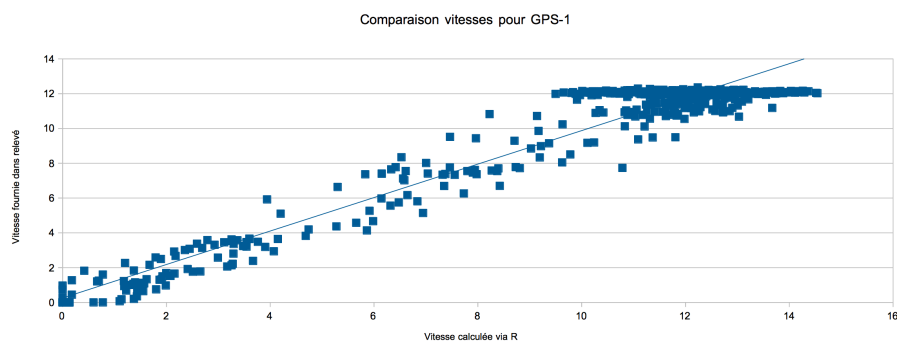


FIGURE 4.5 – **Vitesse GPS-1 en fonction de la vitesse dans le fichier** – Courbe de comparaison entre la vitesse disponible dans le fichier et la vitesse GPS-1 calculée dans R par la méthode Haversine.

Le coefficient de régression pour GPS-1 est de 0,977 et l'écart type est :4,161.

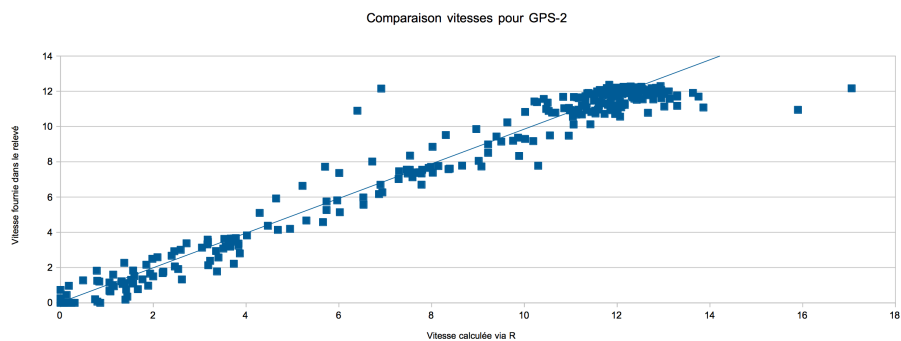


FIGURE 4.6 – **Vitesse GPS-2 en fonction de la vitesse dans le fichier** – Courbe de comparaison entre la vitesse disponible dans le fichier et la vitesse GPS-2 calculée dans R par la méthode Haversine.

Le coefficient de régression pour GPS-2 est de 0,989 et l'écart type est : 4,141.

Les deux valeurs sont donc très proches malgré le fait qu'intuitivement, lors de la visualisation des courbes de vitesse dans le temps, il est constaté une plus grande dispersion des points lors de la plage de vitesse constante pour GPS-1.

En considérant les moyennes de vitesse, il est constaté que pour :

- GPS-1, la moyenne est de 9,778 m/sec ;
- GPS-2, la moyenne est de 9,802 m/sec.

En ce qui concerne la vitesse fournie dans le fichier, la moyenne est de 9,661m/sec (ou 34,78 km/h).

Un examen des coefficients de corrélation pour les vitesses calculées par la méthode "DistGeo" (sans représentation graphique des courbes), permet de constater un coefficient de corrélation de 0,977 pour GPS-1 et un coefficient de corrélation de 0,989 pour GPS-2.

L'écart type est de 4,163 pour GPS-1 et de 4,143 pour GPS-2.

Entre la méthode de calcul Haversine et la méthode de calcul DistGeo, pour GPS-1 :

- coefficient de corrélation : 0,977 et écart type : 4,161 ;
- coefficient de corrélation : 0,977 et écart type : 4,163.

Entre la méthode de calcul Haversine et la méthode de calcul DistGeo, pour GPS-2 :

- coefficient de corrélation : 0,989 et écart type : 4,141 ;
- coefficient de corrélation : 0,989 et écart type : 4,143.

Il en est déduit donc que les valeurs fournies par les deux capteurs GPS sont très proches les unes des autres.

Enfin, concernant les moyennes de vitesse obtenues avec la méthode Dist-Geo, les valeurs sont les suivantes :

- GPS-1, la moyenne est : 9,784 m/sec (ou 35,17km/h) ;
- GPS-2, la moyenne est : 9,809 m/sec (ou 35,28 km/h).

La vitesse fournie dans le fichier, quant à elle, indique toujours une moyenne de 9,661 m/sec.

Comparaison des deux capteurs GPS sur base d'une partie des valeurs

Constatant que les valeurs moyennes fournies par les deux capteurs GPS sont très proches, il n'est pas possible de déterminer formellement quel capteur GPS sera conservé pour la suite des analyses. Dans les premières constatations ci-avant, les valeurs moyennes sont calculées sur l'ensemble des valeurs du fichier. Or, lors du déplacement du tracteur, il est constaté que la vitesse n'est pas linéaire pendant toute la durée de la mesure. On remarque une accélération de la vitesse en début de courbe et une décélération avant une nouvelle accélération.

Pour vérifier la fiabilité, la partie de la courbe pour laquelle la vitesse est +/- constante sera sélectionnée. Cette vitesse constante, dans le fichier, est disponible entre les lignes 139 et 561.

Elle sera sélectionnée dans l'hypothèse que cette plage de données puisse indiquer une plus forte différence dans l'analyse des coefficients de corrélation et d'écart type pour les deux GPS.

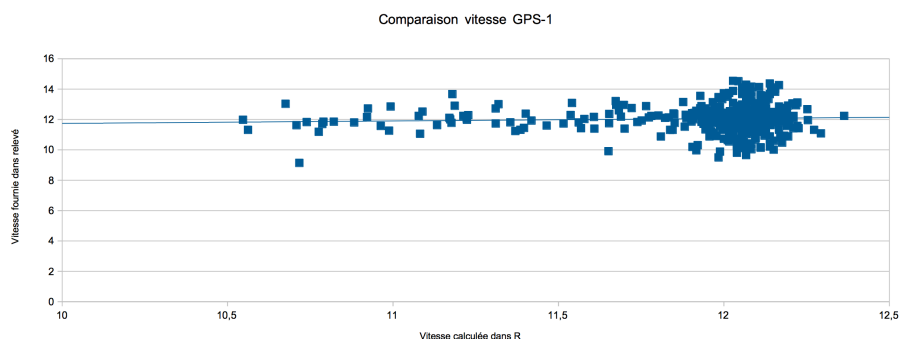


FIGURE 4.7 – **Vitesse GPS-1 en fonction de la vitesse dans le fichier**
– Courbe de comparaison partielle entre la vitesse disponible dans le fichier et la vitesse GPS-1 calculée dans R par la méthode Haversine.

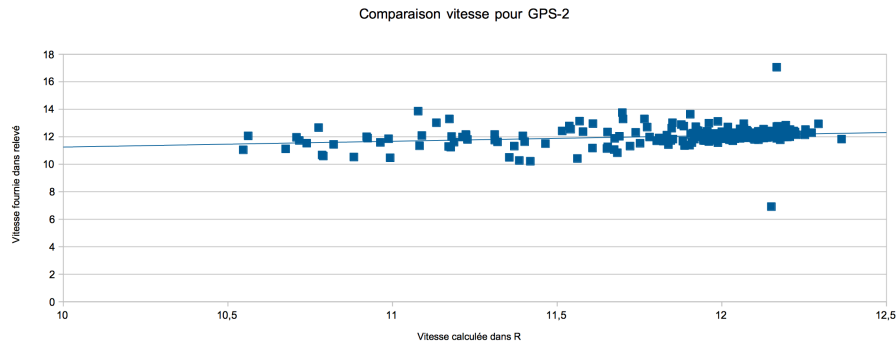


FIGURE 4.8 – **Vitesse GPS-2 en fonction de la vitesse dans le fichier** – Courbe de comparaison partielle entre la vitesse disponible dans le fichier et la vitesse GPS-2 calculée dans R par la méthode Haversine.

Dans ce relevé partiel, le coefficient de corrélation tombe à 0,059 pour le GPS-1 et tombe à 0,250 pour le GPS-2. Le point à retenir de cette comparaison est que les deux capteurs, pour un calcul des mêmes valeurs, ne fournissent pas des valeurs similaires durant toute la période des relevés.

L'écart type calculé est 0,652 pour le GPS-1 et 0,445 pour le GPS-2.

Sur base de ces valeurs, le capteur GPS-2 sera choisi car il a un meilleur coefficient de corrélation que le GPS-1.

Etant donné que les valeurs sont très proches en termes de coefficient de corrélation et d'écart type, tant pour la mesure effectuée sur le jeu complet de données que sur le jeu de données partiel. Cela ne permet pas de constater avec certitude que l'un des deux GPS offre moins de précision que l'autre, malgré une visualisation naïve des données représentées par les graphiques 4.1, 4.2, 4.3 et 4.4.

Il est à souligner également que le coefficient de corrélation tombe à une faible valeur lorsque l'on choisit de ne consulter qu'une partie de la courbe.

4.2.3 Vérification des vitesses calculées pour le deuxième relevé

Les graphiques suivants présentent les relevés des différents GPS pour le deuxième relevé. Il s'agit de graphiques présentant la vitesse du tracteur en fonction de la durée du déplacement.

Relevé sur terrain agricole pour GPS-1

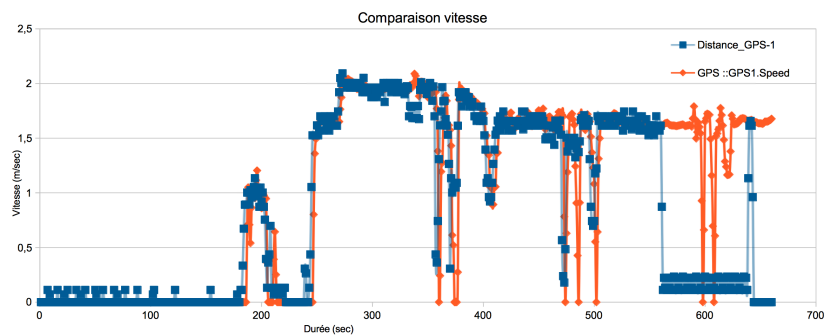


FIGURE 4.9 – **Vitesse calculée suivant la durée du trajet** – Vitesse du tracteur dans la durée pour le 2ème fichier et le capteur GPS-1 calculé par la méthode Haversine.

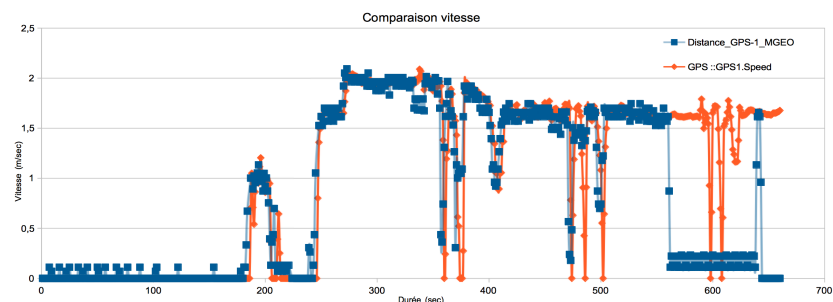


FIGURE 4.10 – **Vitesse calculée suivant la durée du trajet** – Vitesse du tracteur dans la durée pour le 2ème fichier et le capteur GPS-1 calculé par la méthode DistGeo.

Relevé sur terrain agricole pour GPS-2

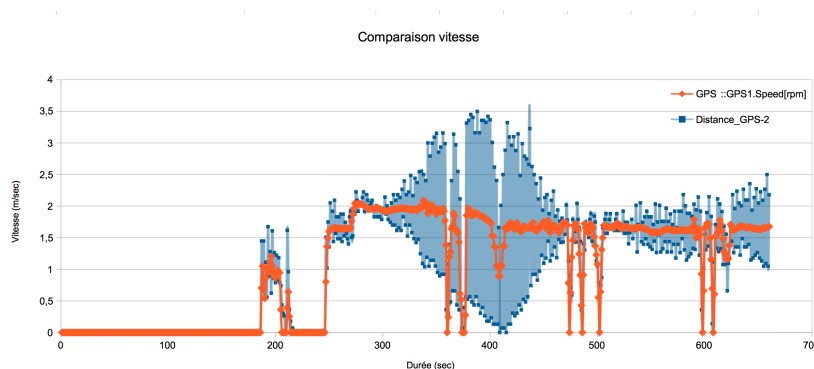


FIGURE 4.11 – **Vitesse** calculée suivant la durée du trajet – Vitesse du tracteur dans la durée pour le 2ème fichier et le capteur GPS-2 calculé par la méthode Haversine.

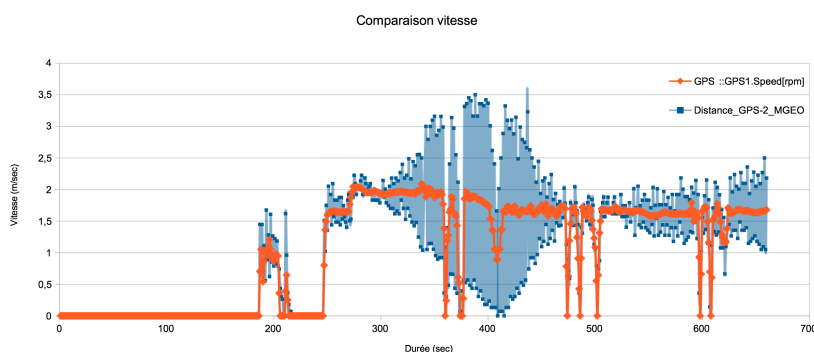


FIGURE 4.12 – **Vitesse** calculée suivant la durée du trajet – Vitesse du tracteur dans la durée pour le 2ème fichier et le capteur GPS-2 calculé par la méthode DistGeo.

Constatations

Les différentes chutes de vitesse dans le tracé des vitesses calculées préalablement sont dues à la manoeuvre de demi-tour du tracteur en fin de ligne droite du champ. Ces chutes de vitesse sont très distinctement

constatées sur les tracés en orange.

Le fait qu'il y ait plusieurs chutes dans un intervalle de temps assez rapproché correspond au fait que l'agriculteur doit effectuer plusieurs manoeuvres pour permettre au tracteur de faire demi-tour.

Plusieurs constatations sont faites grâce à la représentation graphique des jeux de données obtenus :

- comme pour le premier relevé, la courbe de vitesse n'est pas constante dans le temps, et pour ce deuxième relevé, les portions de courbes pour lesquelles la vitesse chute sont certainement liées aux manoeuvres de demi-tour du tracteur ;
- ensuite, il est constaté que la courbe de GPS-1 fournit des points moins dispersés que la courbe de GPS-2 ;
- la vitesse moyenne est moins élevée pour ce relevé que pour le relevé du premier fichier. Cela s'explique par le travail du tracteur dans un champ ;
- enfin, pour chacun des deux relevés GPS, la différence entre la méthode de calcul "Harversine" et la méthode de calcul "DistGeo" ne fournit pas de différence notable.

Les valeurs fournies par le GPS-1 et par le GPS-2 sont vérifiées. Afin de faciliter les analyses ultérieures, il sera intéressant de ne tenir compte que d'un seul des deux capteurs GPS.

Comme pour le premier relevé, une comparaison des valeurs disponibles dans les deux GPS est faite pour tenter de déterminer quel capteur est le plus fiable.

Pour ce faire, la valeur de vitesse fournie dans le fichier de base est tout d'abord comparée avec la valeur calculée par la méthode "Harversine" du script "R" afin d'évaluer la précision.

Le choix de la régression se porte sur une régression linéaire puisqu'il est supposé qu'il existe une relation linéaire entre les deux mesures. Il est supposé que la valeur de vitesse augmente de la même manière pour les données disponibles dans le fichier et pour les valeurs calculées. La variable explicative est la valeur disponible dans le fichier.

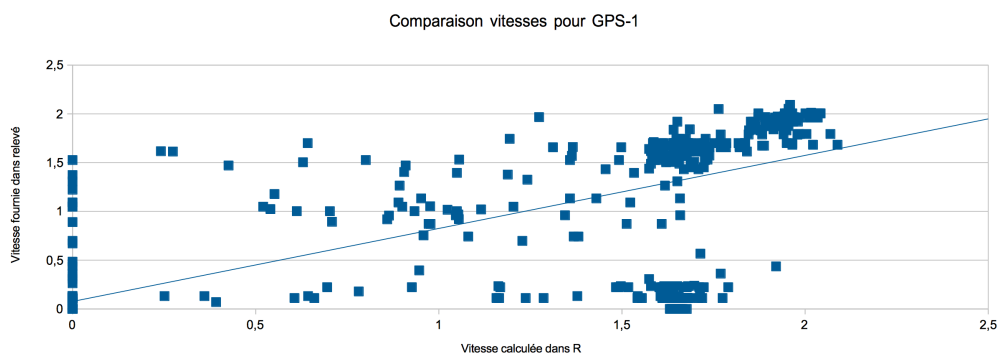


FIGURE 4.13 – **Vitesse GPS-1 en fonction de la vitesse dans le fichier**
– Courbe de comparaison entre la vitesse disponible dans le fichier et la vitesse GPS-1 calculée dans R par la méthode Haversine.

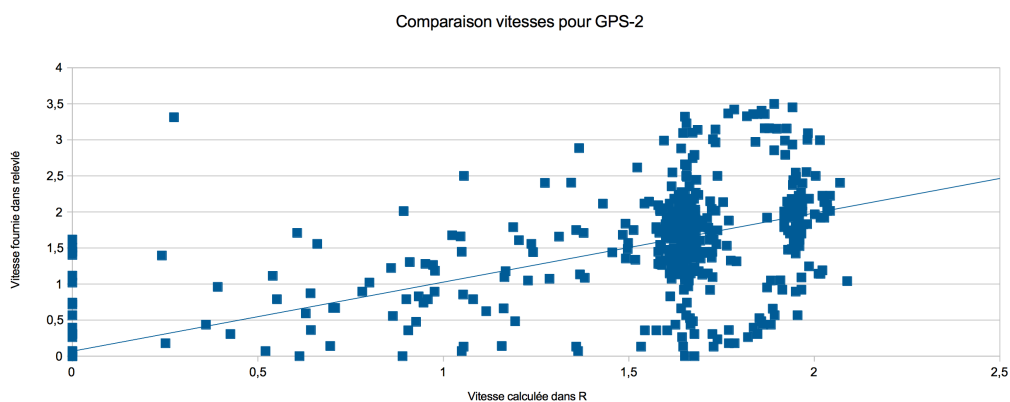


FIGURE 4.14 – **Vitesse GPS-2 en fonction de la vitesse dans le fichier**
– Courbe de comparaison entre la vitesse disponible dans le fichier et la vitesse GPS-2 calculée dans R par la méthode Haversine.

Le coefficient de régression pour GPS-1 est de 0,754 et l'écart type est : 0,820.

Le coefficient de régression pour GPS-2 est de 0,811 et l'écart type est : 0,895.

Dans ce deuxième relevé, la corrélation est plus forte pour le GPS-2. Cette plus forte corrélation est certainement induite par le comportement des valeurs dans GPS-1 pour la plage de données allant +/- la seconde 550 jusqu'à la fin de la mesure et pour laquelle les valeurs du GPS ne suivent plus les valeurs fournies dans le fichier.

Ceci met en évidence le fait que le GPS peut également avoir un problème technique ne permettant plus de délivrer les bonnes valeurs.

En considérant les moyennes de vitesse, il est constaté que pour :

- GPS-1, la moyenne est de 0,858 m/sec (ou 3,092 km/h) ;
- GPS-2, la moyenne est de 1,068 m/sec (ou 3,84 km/h).

En ce qui concerne la vitesse fournie dans le fichier, la moyenne est de 1,044 m/sec (3,74 km/h).

Pour ce deuxième relevé, le capteur GPS-1, dont le coefficient de corrélation est plus important, sera utilisé.

4.3 Comparaison de la vitesse avec le nombre de tours/minutes

4.3.1 Premier relevé avec mesures du capteur GPS-2

La vitesse du tracteur, calculée via les différentes positions GPS, est évaluée avec le nombre de tours/minutes du moteur.

Le nombre de tours/minutes sera évalué pour chacun des deux fichiers analysés. Comme présenté précédemment, le capteur GPS-1 correspond au premier relevé et GPS-2 correspond au deuxième relevé.

Est-il possible que cette vitesse soit directement dépendante du nombre de tours/minutes ?

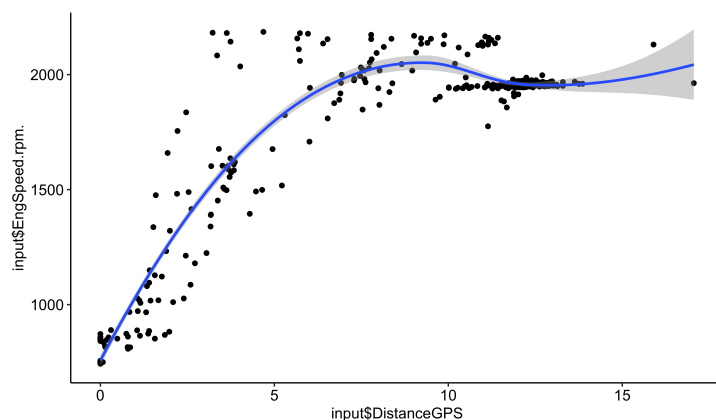


FIGURE 4.15 – **Vitesse calculée en fonction du nombre de tours/minutes** – Vitesse du tracteur via le capteur GPS-2 en fonction du nombre de tours/minutes du moteur.

Le coefficient de corrélation pour ce premier diagramme est de : 0.876

4.3.2 Deuxième relevé avec mesures du capteur GPS-1

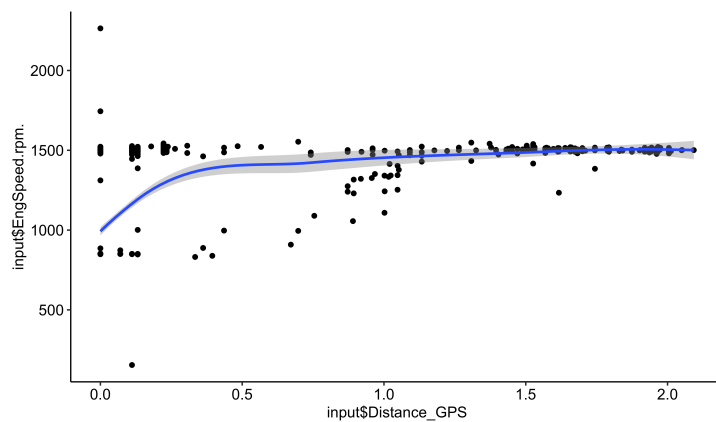


FIGURE 4.16 – **Vitesse calculée en fonction du nombre de tours/minutes** – Vitesse du tracteur via le capteur GPS-2 en fonction du nombre de tours/minutes du moteur.

Le coefficient de corrélation pour ce premier diagramme est de : 0.671.

ActualEngPercentTorque[%] pour le premier fichier : Moyenne 36,76%

ActualEngPercentTorque[%] pour le premier fichier : Moyenne 40,96%
lorsque la vitesse du tracteur est à 12 m/sec

ActualEngPercentTorque[%] pour le premier fichier : Moyenne 26,28%
lorsque la vitesse du tracteur est inférieure à 12 m/sec.

La base de choix de la vitesse de 12 m/sec est déterminée par la vitesse moyenne lorsque le tracteur est à une vitesse de déplacement constante. Ce déplacement constant est constaté entre les secondes 139 et 561.

ActualEngPercentTorque[%] pour le deuxième fichier : 47,34%

Il existe un grand nombre de variations dans la vitesse du tracteur lorsque celui-ci est en travail de champ et une moins bonne corrélation entre le régime moteur et la vitesse. Il peut être supposé que cette moins bonne corrélation est liée à la nature du sol sur lequel le tracteur se déplace, mais également liée au fait que le tracteur est en action de travail de champ (dans le cas présent, une action de pulvérisation).

Enfin, il peut en être déduit que si, sur route, la corrélation semble importante, signifiant que, sans contraintes complémentaires, les données fournies directement par le tracteur peuvent être manipulées sans corrections, ce n'est plus le cas lorsque celui-ci est en travail au champ. Etant donné que c'est en travail au champ que les données sont les plus intéressantes à consulter, il sera nécessaire d'appliquer une méthode de régression avant de pouvoir manipuler les données.

Chapitre 5

Algorithmes de correction

Comme constaté dans les chapitres précédents, les valeurs fournies par le Bus-CAN peuvent ne pas être exploitables en l'état. Il est nécessaire de les vérifier et, au besoin, les corriger avant exploitation :

- soit en les comparant avec des valeurs mesurées sur un banc d'essai ;
- soit en appliquant des formules de régression.

Il reste néanmoins un grand nombre de difficultés à surmonter pour obtenir des données exploitables d'un tracteur agricole.

Tout d'abord, de par l'utilisation du tracteur, les contraintes pouvant générer des erreurs dans les données sont nombreuses et doivent être prises en compte lors d'une analyse des données. Ces contraintes sont :

- la nature du sol sur lequel se déplace le tracteur ;
- la nature de l'opération en cours à réaliser par le tracteur ;
- les conditions environnementales lors de l'utilisation du tracteur ;
- les outils complémentaires ajoutés au tracteur pour effectuer son travail ;
- ...

Plusieurs approches sont possibles pour tenter d'appliquer les corrections nécessaires :

- la sélection d'une plage de données parmi le volume complet de données disponibles ;
- une méthode statistique classique utilisant une comparaison avec des valeurs mesurées sur un banc d'essai ;

- une approche empirique en fonction de l'analyse à effectuer ;
- l'application d'une méthode de machine learning pour permettre un traitement efficace sur l'ensemble des données ;
- ...

La sélection d'une plage de données est couramment proposée comme constaté à la lecture des études dans les chapitres précédents. Néanmoins, elle offre le désavantage de ne pas consulter l'ensemble du jeu de données. Il est supposé que les données non considérées dans les plages utiles seront également des données inutiles à stocker. En considérant la notion de stockage et de maintenance de ces données dans un système informatique, il pourrait être opportun de se poser la question du volume de données non utiles qui reste stocké dans les bases de données.

Il peut être également opportun de réfléchir à un processus de Machine Learning pour exploiter les données. Comme constaté à la lecture des analyses faites plus haut, les valeurs de vitesse, de notre analyse, doivent être classifiées par type de travail (travail au champs, demi-tours). Comparativement à une méthode statistique, l'ensemble des données, utilisées dans un processus de Machine Learning, peut être considéré et il n'est pas nécessaire d'effectuer ces classifications des données. Dans le processus de Machine Learning pour valider des données, celles qui seront disponibles pour l'apprentissage, devront être de bonne qualité.

Il est à supposer que cette qualité (définie par plusieurs attributs tels que le type, la plage correctement définie, l'unicité, le format, la validation préalable) est requise pour permettre un bon entraînement de la machine à faire face aux données à vérifier. Ces données d'entraînement seront donc relevées par un processus clair et rigoureux (sur un banc d'essai, avec plusieurs tests distincts, suivant un cahier des charges défini, etc).

Il existe encore peu d'études concernant le processus de Machine Learning à propos des données disponibles par le Bus-CAN d'un tracteur agricole pour traitement a posteriori, néanmoins le sujet a été abordé et quelques éléments d'analyse sont proposés dans ce document.

5.1 Réseau de neurones

La première étude sur le sujet [32] cherche à définir s'il est possible d'appliquer un traitement par réseau de neurones à propos de la détection de pannes sur un véhicule à moteur équipé du protocole SAE J1939.

Le diagnostic de panne dépend fortement de l'expérience humaine pour être efficace. Il s'agit principalement de récupérer les données de diagnostic pour qu'elles soient analysées par quelqu'un qui pourra ensuite déterminer le type de problème. Pour améliorer le diagnostic, un réseau de neurones RBF (Radial Basis Function) est utilisé. Le réseau de neurones RBF est un réseau à deux couches avec une couche cachée.

La couche d'entrée se compose des nœuds de départ. Le nombre d'unités de couches cachées dépend du problème à résoudre. La transformation de la couche d'entrée vers la couche masquée est non linéaire et la transformation de la couche masquée vers la zone de sortie est linéaire. Les nœuds de la couche cachée se composent de fonctions telles que la fonction gaussienne. Le nœud de couche de sortie est généralement une fonction linéaire simple.

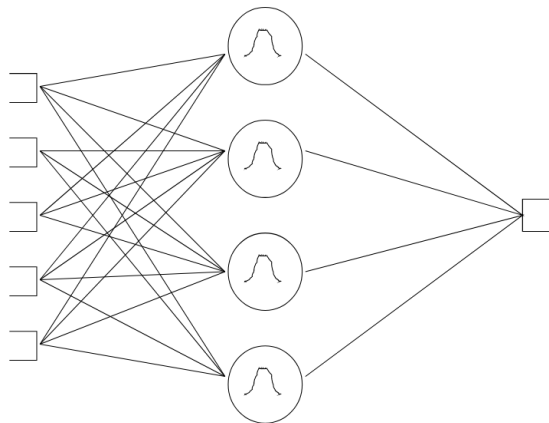


FIGURE 5.1 – Réseau de neurones RBF. Figure reproduite depuis [32], figure III.

Dans cette étude, 7 points d'entrée sont définis. Une entrée par type de

panne associée à un capteur :

- x1 : pression mesurée dans la rampe d'injection ;
- x2 : vitesse du véhicule ;
- x3 : pression d'huile ;
- x4 : température du liquide de refroidissement ;
- x5 : vitesse de rotation ;
- x6 : température de sortie SCR ;
- x7 : température d'entrée SCR.

Un algorithme d'optimisation est également utilisé. Il s'agit d'un algorithme à essaim de particules (Particle Swarm Optimization - PSO). Cette méthode d'optimisation méta-heuristique stochastique permet de résoudre des questions complexes via des processus algorithmiques simples. Néanmoins, étant donné que le processus peut parcourir un réseau de neurones un grand nombre de fois, le principal problème de cette méthode réside dans son temps d'exécution. Cette méthode n'est pas optimale pour de grands volumes de données.

Les méthodes méta-heuristiques, sont des méthodes d'optimisation permettant d'obtenir une valeur proche de la meilleure solution dans un délai raisonnable d'exécution. Elles ont pour but de résoudre des problèmes dans différents domaines sans avoir à modifier l'algorithme de base.

Des groupes de données (56) ont été sélectionnés comme données d'apprentissage et le reste des données disponibles a été considéré comme les données à vérifier. Les entrées sont au nombre de 7 suivant les descriptions de types de pannes faites préalablement et 25 neurones sont positionnés dans la couche cachée. La conclusion de l'étude constate notamment que l'algorithme offre une certaine efficacité pour la détection des pannes mais qu'il pourrait être amélioré par la suite pour être efficace à 100% en un temps plus court.

L'étude [33] cherche à évaluer la mise en oeuvre d'un réseau de neurones pour mesurer les capteurs d'un tracteur agricole. Les capteurs choisis sont la température des gaz d'échappement et, alternativement, la température d'huile moteur. Le constat de départ est que l'Italie souffre d'un parc de véhicules agricoles assez vieux pour lequel beaucoup de tracteurs ne sont pas encore équipés du réseau Bus-CAN. Le projet prévoit de mesurer la température d'échappement à l'aide de thermocouple. Un thermocouple de type K

(thermocouple d'utilisation générale) est prévu en sortie de l'échappement et un autre a été plongé dans le carter d'huile. La température des gaz d'échappement est économiquement intéressante à mesurer et assez précise pour calculer simplement le couple moteur et le régime moteur.

$$Tem = f1(M; n) \iff M = f2(Tem; n) \quad (5.1)$$

(Équation 1, page 2 dans étude [33]).

Où :

- Tem : est la température relevée en sortie d'échappement ;
- M : est le couple moteur ;
- n : est le régime moteur.

Le but de cette analyse est de trouver une méthode algorithmique permettant de décrire les corrélations entre la température et la vitesse du moteur. L'analyse de réseau de neurones artificiels a été proposée pour évaluer numériquement la valeur de certaines données (couple moteur et BSFC - Brake Specific Fuel Consumption) en ne considérant que les données de référence décrites ci-dessus, combinées avec le régime moteur. Ces valeurs de référence ont d'abord fait l'objet d'un relevé via des mesures en banc d'essai. Quatre tracteurs sont sélectionnés pour les mesures de référence pour un total de 219 relevés.

Un réseau de neurones, aussi appelé méthode connexionniste, adapte son comportement en fonction du "poids" des valeurs, des interconnexions de neurones et de la fonction de transfert entre les neurones. Il est amélioré par le biais de données d'apprentissage basées sur des valeurs empiriques.

L'analyse prévoit une première phase d'apprentissage durant laquelle le poids de chaque neurone est modifié de manière itérative. Cette modification intervient par le biais d'un algorithme de rétro-propagation d'erreurs. Le but étant de faire correspondre à un ensemble de couple de valeurs entrée/sortie. L'apprentissage se termine lorsque l'algorithme est capable de minimiser les erreurs en sortie de réseau de neurones.

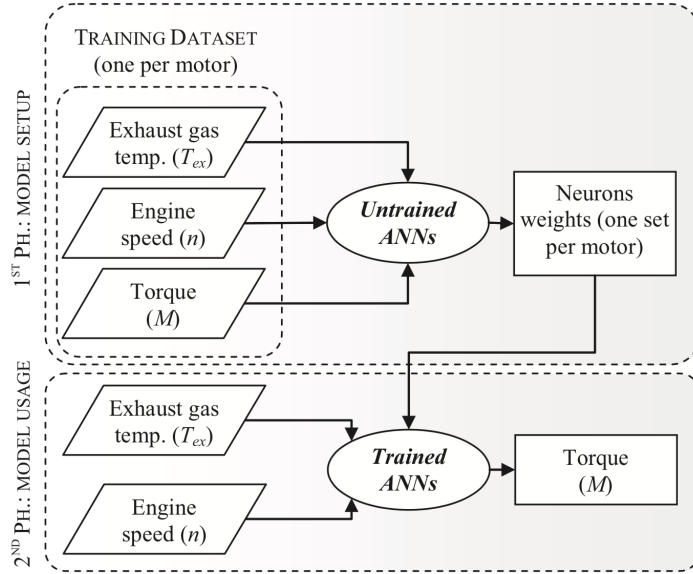


FIGURE 5.2 – Figure reproduite de l'étude [33] (fig.4, page 5), de l'entraînement du réseau de neurones.

Dans cette étude, le choix de la configuration du réseau a été réalisé de manière empirique. Il a fait l'objet d'investigations. Le choix est fait de proposer, pour une modélisation de 4 ensembles de réseaux différents, la configuration suivante :

- une couche d'entrée avec deux noeuds (chaque noeud correspondant à un type de variables) ;
- une couche puis deux couches cachées avec 26 et/ou 10 neurones ;
- une couche de sortie.

De plus, différentes dispositions de réseaux de neurones ainsi que différentes fonctions de transferts ont été testées. Finalement, le choix s'est porté sur l'utilisation de deux couches cachées. Le processus d'apprentissage a le choix d'exclure certains noeuds. Plusieurs jeux de données distincts, issus des différents tracteurs sélectionnés pour l'étude, ont été assignés suivant deux sous-ensemble de données, de manière aléatoire, à plusieurs configurations de réseaux de neurones comme décrits ci-avant. Dix configurations de réseaux de neurones ont été formées pour chaque tracteur sélectionné dans l'étude.

Au total, 160 analyses ont été faites. Le temps de convergence était fixé à 10000 itérations.

Les résultats donnent un temps d'entraînement moyen de 2,87 secondes. Dans le pire des cas, le temps d'entraînement atteint 4,87 secondes. Le temps d'apprentissage est lié :

- à la taille de l'ensemble de données ;
- au nombre de neurones dans les couches cachées ;
- au nombre de connexions initiales.

Le R^2 des trois modèles de régression linéaire pour les prévisions de temps est toujours supérieur à 0,941, ce qui suppose qu'ils peuvent être utilisés pour dimensionner le réseau de neurones.

Cette étude prouve que l'utilisation des réseaux de neurones améliore la précision des pannes grâce à une meilleure prédiction du couple moteur et qu'elle est pleinement justifiée. Il est constaté que suivant les caractéristiques internes du tracteur sélectionnées pour l'analyse, certains types de réseaux neuronaux peuvent donner de moins bons résultats de prédictions. Les valeurs de R^2 prédisant une consommation BSFC (Brake Specific Fuel Consumption) diffèrent suivant le type de fonction neuronale utilisée. L'utilisation de la fonction de transfert gaussienne augmente les capacités prédictives. Le choix de la fonction de transfert dans la couche de neurones la plus proche de la sortie a également une influence dans les capacités de prédictions avec des améliorations de prédictions plus importantes pour la consommation BSFC que pour le couple moteur. La température de l'huile moteur ne s'est pas avérée être un bon estimateur de la consommation BSFC ou du couple moteur. Enfin, l'utilisation de réseau de neurones s'est avéré utile pour corrélérer la température et les régimes moteurs avec le couple et la consommation BSFC.

L'étude [34] propose un réseau de neurones BP (Back Propagation Neural Network) pour tenter de détecter des défauts de fonctionnement d'un tracteur, particulièrement lorsque l'on doit faire face à une panne de type "multifault" (contenant plusieurs défauts). Le dispositif Bus-CAN servait à collecter les données de 8 éléments du moteur et un algorithme d'optimisation a été utilisé pour améliorer la structure et le poids du réseau de neurones. Le réseau de neurones BPNN, combiné à la méthode

d'optimisation, est performant en termes de vitesse mais également en termes de généralisation et de précision. Un algorithme d'apprentissage a pour objectif de corriger le poids de la connexion entrante du neurone et le seuil de l'erreur entre la sortie réelle et la sortie attendue de l'algorithme.

L'objectif de cette étude est d'améliorer la vitesse et la précision de détection d'une panne moteur d'un tracteur agricole. Le principe est de pouvoir détecter le défaut au stade premier de la panne pour la résoudre en améliorant les performances du tracteur. 8 capteurs du Bus-CAN ont été utilisés, notamment la vitesse du moteur, la charge du moteur, le flux d'air en entrée du moteur, etc. Plusieurs méthodes d'optimisation ont été testées.

Le réseau de neurones BPNN adopte une architecture en 3 couches : une couche d'entrée, une couche cachée et une couche de sortie. Pendant le processus d'entraînement, la précision d'entraînement du réseau de neurones BPNN est mesuré par l'erreur d'apprentissage RMSE [35] (Root Mean Squared Error).

Pour le réseau de neurones BPNN, il existe une descente de gradient qui dépend des valeurs initiales sélectionnées. Cette descente de gradient est utilisée dans la rétro-propagation, mais cette méthode peut avoir un temps de formation important. Néanmoins, les algorithmes d'optimisation peuvent améliorer cet apprentissage. Cette étude démontre que l'utilisation de réseau de neurones, combiné à des algorithmes d'optimisation, peuvent parfaitement remplir le rôle de détection précoce des pannes d'un moteur de tracteur agricole.

La dernière étude analysée avec utilisation d'un réseau de neurones [36] cherche à déterminer de manière fiable la consommation de carburant d'un tracteur agricole. Ce réseau de neurones est combiné à 6 algorithmes d'entraînement différents. Les 6 algorithmes testés concernent la descente de gradient permettant de trouver la valeur minimale de l'erreur possible.

Dans cette étude, la recherche du nombre de neurones de la couche cachée a été réalisée via la méthode essai/erreur. Le nombre de neurones dans la couche cachée a été choisi en vérifiant les performances globales du réseau de neurones.

Le meilleur algorithme de descente de gradient ainsi que les performances

globales du réseau de neurones ont été comparés sur base de l'échelle d'erreur quadratique moyenne (MSE), de la somme des erreurs au carré (SSE), du coefficient de détermination (R^2) et de la précision des prévisions (PA). Le réseau comportant 2 couches cachées de 10 neurones dans chaque couche a été le réseau comportant le moins d'erreurs et ayant la plus grande prévision de simulation.

Le réseau de neurones à rétro-propagation avec l'algorithme d'entraînement Levenberg-Marquardt est celui qui présente les meilleures performances. L'utilisation du réseau de neurones a été comparé avec des méthodes statistiques traditionnelles. Il s'avère que l'utilisation d'un réseau de neurones est efficace pour prédire une consommation de carburant proche des valeurs réellement capturées, comparativement à d'autres modèles statistiques classiques.

5.2 Machine à vecteurs de support

Pour estimer les valeurs de couple d'un moteur en fonction de la vitesse du moteur [37], les approches statistiques classiques ont montré leurs limites car pour beaucoup de cas, pour des estimations traditionnelles basées sur des modèles dynamiques, il est nécessaire d'avoir des valeurs précises pour les éléments du moteur à utiliser. La connaissance précise de ces valeurs peut parfois être très coûteuse.

Une méthode efficace d'estimation fait appel à l'apprentissage automatique. Dans cette étude, le choix est fait d'utiliser des machines à vecteurs de support (SVM - Support Vector Machine) qui optimisent les problèmes de classification et d'estimation de fonctions non linéaires. Les capacités de régression des SVM ont montré un meilleur potentiel que les méthodes classiques pour généraliser une problématique avec une précision valable, tout en maintenant une grande simplicité dans son processus d'estimation.

Les SVM (Support Vector Machine) sont des types de méthodologie de machine learning. L'objectif est d'apprendre via des données d'apprentissage basées sur des éléments d'entrée en fonction de la sortie souhaitée. Les SVM fonctionnent dans un espace, communément appelé "espace caractéristique",

qui est une transformation des données fournies en entrée. Grâce à cette transformation non linéaire, la solution permettra de faire correspondre les valeurs d'entrée avec les valeurs de sortie de manière semblable à une méthodologie de correction de courbe. Pour chaque vecteur de support, il existe un produit entre le poids et une fonction de mapping non linéaire.

$$f(X) = (W, \Phi(X)) + b \quad (5.2)$$

(Équation 2, page 6 dans étude [37]).

Où :

- W : le poids de la matrice ;
- $\Phi(X)$: le vecteur de la fonction de mapping non linéaire entre l'entrée et la sortie ;
- b : représente la valeur du biais.

Contrairement aux réseaux de neurones, pour lesquels il est nécessaire de définir la bonne architecture, dans le cas présent, c'est la sélection d'un noyau approprié qui prévaut. Une des fonctions de noyau possible est la fonction à base radiale (RBF). Les données du moteur sont préparées en données d'entraînement, le programme utilisé pour l'entraînement SVM est SVM-Light [38]. Le résultat est la précision de prédiction du couple moteur par l'utilisation de la SVM à partir des données de couple indiquées et de la pression dans les cylindres.

Il est nécessaire, pour un bon entraînement de la SVM de faire une sélection correcte des données d'apprentissage. Il n'est pas non plus souhaitable que l'apprentissage par SVM se fasse sur l'ensemble des données disponibles pour chaque jeu de données. Un algorithme a été élaboré pour ne conserver que les données d'entraînement les plus pertinentes.

Le SVM peut prédire la variabilité du couple efficacement en ne produisant que de très petites erreurs. Néanmoins, il existe des plages de valeurs pour lesquelles la SVM produira des erreurs plus importantes. L'utilisation de la technologie SVM permet de reconstruire efficacement la valeur réelle du couple moteur à différents régimes moteur. L'avantage de la technique réside dans la fourniture de valeurs de sortie efficaces tout en utilisant un petit ensemble de données en entrée.

Chapitre 6

Conclusion

Les données Bus-CAN d'un tracteur en travail au champ ne peuvent être exploitées en l'état. Les contraintes liées au type de travail ou à la nature du sol influencent le comportement du véhicule et peut introduire des erreurs dans les relevés.

Comme constaté dans les diverses études menées sur le sujet, une méthode consiste à utiliser des équations empiriques avec un grand nombre de données pour essayer de répondre à la question de la précision des données. Néanmoins cette méthode, bien que fiable, peut devenir très fastidieuse à mettre en oeuvre, tant par le nombre de relevés à réaliser que par les résultats d'équations à produire. De plus, certaines études suggèrent également de réaliser ces équations empiriques sur un grand nombre de marques et modèles distincts de tracteurs.

La constatation est faite également que, lors de nos propres relevés, les conditions d'utilisation d'un tracteur pour un travail agricole peuvent varier fortement avec énormément de rejets de valeurs qui ne seraient pas considérées comme étant des valeurs en pleine charge.

Cela signifie donc qu'il est nécessaire de sélectionner les plages de valeurs qui sont opportunes pour une analyse des données et que ces sélections doivent se répéter :

- pour des types de travail de champ distincts ;
- pour des marques et modèles de tracteurs distincts ;
- dans des conditions d'utilisation multiples ;

— ...

Concernant les relevés de vitesse par le biais des positions GPS

Il en est déduit que la moins bonne précision de la position GPS (et donc le calcul de la vitesse instantanée) est directement liée aux variations de la vitesse du tracteur. Il existe plus d'imprécisions dans les relevés qui contiennent des variations régulières de vitesse.

L'approche d'une analyse des données par le biais d'une méthode d'apprentissage par machine learning permet d'éviter de manipuler manuellement les équations, la lecture des résultats et l'interprétation de ceux-ci. Elle permet de ne pas effectuer de classification des données, entraînant un rejet d'une partie de celles-ci. Néanmoins, il est nécessaire de bien choisir la structure du réseau de neurones (nombre de points d'entrée, nombre de couches cachées et nombre de neurones dans la couche cachée) ainsi que les fonctions de transferts en rapport au type de problème à résoudre. Les algorithmes d'entraînement liés aux réseaux de neurones jouent également un rôle essentiel dans la précision des résultats et dans le temps de traitement global.

Comme constaté dans les études analysant les corrections de valeurs via des réseaux de neurones, ceux-ci montrent une certaine efficacité pour la résolution de problèmes d'estimation statistique liés aux valeurs fournies par les tracteurs agricoles pour être utilisés dans le domaine de l'agriculture de précision. Néanmoins, la problématique persistante concerne le choix de la configuration du réseau de neurones.

Il est supposé que, pour un problème défini comme celui de vérifier la fiabilité du capteur de régime moteur, des solutions empiriques de réseaux de neurones peuvent être proposés pour tenter d'établir le meilleur rendement. Le rendement peut être évalué sur base de deux critères : la précision des données en sortie et la vitesse d'exécution du réseau de neurones. Les SVM, quant à elles, nécessitent toujours de faire une classification des données pour permettre un résultat valable.

Il existe plusieurs types d'algorithmes pour le Machine Learning. Bien que quelques-uns soient décrits dans le chapitre précédent, d'autres peuvent encore être décrits. Le choix du bon algorithme doit se faire de manière éclairée en fonction du type de problème à résoudre.

Ces algorithmes sont les suivants :

Algorithme supervisé de régression linéaire

Parmi les plus utilisés, il permet d'établir des corrélations simples entre deux variables dans un set de données. Modèle simple à utiliser, il permet de montrer comment le changement intervenant dans une variable peut avoir un impact sur une autre variable.

Algorithme supervisé de Machine à vecteurs de support

Ce sont des algorithmes qui séparent les données en classes distinctes. Les classes peuvent être définies comme étant des données ayant des caractéristiques communes. Pendant la phase d'apprentissage, l'algorithme cherchera une ligne de séparation entre les données mais cherchera également à maximiser les marges entre les données les plus proches de la ligne de séparation et celles-ci. Mais il est également possible, comme constaté avant, d'utiliser l'algorithme pour des données non-linéaires.

Algorithme supervisé d'arbre de décision

Cet algorithme permet une représentation graphique des données sous forme de branches pour donner les résultats des diverses actions. Il prédit la réponse en fonction des décisions prises, mais reste peu efficace lorsque les volumes de données augmentent.

Algorithme non supervisé - algorithme A priori

Algorithme cherchant les affinités entre deux éléments d'un jeu de données pour identifier s'il existe une corrélation entre ces éléments.

Algorithme non supervisé de répartition en K-moyennes (K-means)

Algorithme s'appuyant sur une méthode itérative pour trier les données en classes basées sur des caractéristiques similaires. Il a la réputation d'être précis tout en ayant un temps de traitement relativement faible.

Algorithme semi supervisé de réseaux antagonistes génératifs

Deux réseaux de neurones sont placés en compétition pour trouver la meilleure solution. Un des deux réseaux, dit générateur, effectue l'analyse des données en entrée pour donner le meilleur résultat possible, alors que le second réseau, dit discriminateur, s'appuie sur les sorties du premier réseau pour en améliorer les résultats.

Algorithme semi supervisé de Classificateur bayésien naïf

Il s'appuie sur le théorème de Bayes qui est fondé sur les probabilités conditionnelles. Il permet de reconnaître des classes d'objets sur des jeux de données étiquetés. Ensuite l'algorithme est appliqué à des jeux de données non étiquetés pour rechercher les étiquettes.

Apprentissage par renforcement - Q-Learning

Algorithmes qui cherchent à trouver la meilleure méthode pour atteindre l'objectif défini. Ils tentent d'effectuer, de manière itérative, le plus grand nombre d'actions possibles sans connaissance préalable de l'environnement des données. Il est souvent associé à des modèles de Deep Learning.

Apprentissage par renforcement - Algorithme basé sur un modèle (model-based)

A l'inverse de la méthode précédente, cet algorithme est limité dans les actions possibles, ce qui lui donne une plus grande performance car il est formé avec des données spécifiques suivant l'environnement des données.

Il existe plusieurs possibilités d'algorithme permettant de résoudre les problèmes d'apprentissage. Il est donc nécessaire de sélectionner le bon algorithme en fonction du type de données à analyser, mais également du type de problème à résoudre.

A titre personnel

Du point de vue de mon apprentissage personnel et en regard des besoins informatiques liés à mon travail au CRA-W, la conclusion de cette étude concerne l'apport non négligeable de nouvelles connaissances. Ce travail m'a permis de découvrir les bases du langage "R", de pouvoir apporter

une réflexion mathématique en préalable à une analyse informatique pour essayer de définir des problèmes complexes.

Cet apprentissage complémentaire est un atout pour mon emploi actuel qui est dédié à des développements informatiques dans le domaine de la recherche agronomique. De plus, il est régulièrement demandé aux informaticiens, dans le cadre de leur travail, de traduire dans une application complète des éléments de travaux réalisés dans le langage "R" par les équipes de recherches.

Donc, le fait de mieux comprendre les mécanismes mathématiques de prédictions, de statistiques et de préparation à des processus de "machine learning", de même que de pouvoir interpréter du code "R", et principalement les concepts de recherche qui sont écrits dans les scripts, favorise grandement les échanges qui peuvent y avoir entre les chercheurs et les informaticiens pour, in fine, obtenir des produits informatiques fiables et qui répondent aux besoins souhaités.

Egalement, une réflexion globale sur le stockage et la gestion des données semble importante, au regard des conclusions de cette analyse. C'est également le rôle de l'informaticien de réfléchir à cette problématique tout en permettant de conserver des données consistantes pour la recherche.

Bibliographie

- [1] Perspectives Agricoles. « *La révolution numérique : comment l'agriculture s'approprie-t-elle le digital ?* ». Flers CEDEX. Mise à jour le 01 mars 2021. [Consulté le 10 mars 2021]. Consultable à l'adresse : <https://www.perspectives-agricoles.com/la-revolution-numerique-comment-l-agriculture-s-approprie-t-elle-le-digital--@/view-3653-arvarticlepa.html?>.
- [2] Corentin Leroux. « *Standards et échanges de données dans le numérique agricole* ». Montpellier. Mise à jour le 17 février 2021. [Consulté le 2 mars 2021]. Consultable à l'adresse : <https://www.aspexit.com/standards-et-echanges-de-donnees-dans-le-numerique-agricole/>.
- [3] G. Grenier. « *Bus-CAN sur machines agricoles : les technologies de l'information au service de l'agriculture de précision et de la traçabilité* ». Ingénieries - E A T, IRSTEA édition 2001, p. 67 - p. 76.
- [4] S. Janković, D. Kleut, I. Blagojević, D. Stamenković, G. Vorotović. « *Application of Vehicle's CAN based Network in Transmission Service Load Data acquisition* ». Tehnicki Vjesnik 19, 2012, p. 201 - 210.
- [5] SAE International. « *SAE J1939 Standards Collection on the Web : Content* ». [Consulté le 15 novembre 2020]. Consultable à l'adresse : <https://www.sae.org/standardsdev/groundvehicle/j1939a.htm>.
- [6] Isobus. « *ISOBUS Data Dictionary according to ISO 11783-11* ». [Consulté le 15 novembre 2020]. Consultable à l'adresse : <https://www.isobus.net/isobus/>.

- [7] Organisation Internationale de Normalisation. « *ISO 11783-1 :2017(fr)* ». Consulté le 10 avril 2021]. Consultable à l'adresse : <https://www.iso.org/obp/ui/#iso:std:iso:11783:-1:ed-2:v1:fr>.
- [8] Nannan Lin B.S. « *Prediction of Wheeled Tractor Fuel Use Rate Distributions from Bus-CAN Data for Agricultural Field Operations* ». Thesis. Graduate Program in Food, Agricultural and Biological Engineering. The Ohio State University, 2014.
- [9] Samuel E. Marx. « *Controller Area Network (CAN) Bus J1939 Data Acquisition Methods and Parameter Accuracy Assessment Using Nebraska Tractor Test Laboratory Data* ». Spring 5-2015, University of Nebraska - Lincoln, United States, 2015.
- [10] S. Lacour, C. Burgun, C. Perilhon, G. Descombes, V. Doyen. « *A model to assess tractor operational efficiency from bench test data* ». ScienceDirect. Journal of Terramechanics 54 (2014) 1–18. F-69360 Solaize, France, April 2014.
- [11] Jacek Wasilewski, Andrzej Kuranc, Joanna Szyszlak-Barglowicz, Monika Stoma, Tomasz Slowik, Dalibor Barta. « *Assesement of efficiency of an agricultural tractor engine for different rotational speeds* ». IX International Scientific Symposium. "Farm Machinery and Processes Management in Sustainable Agriculture", Lublin, Poland, 2017.
- [12] Dimitrios Rimpas, Andreas Papadakis, Maria Samarakou. « *OBD-II sensor diagnostics for monitoring vehicle operation and consumption* ». Tmrees, EURACA, 04 to 06 September 2019, Athens, Greece.
- [13] J. Čupera, P. Sedlák. « *Design and verification of engine power caclulation model using the data of a digital bus built into an agricultural tractor* ». Acta univ. agric. et silvic. Mendel. Brun., 2011, LIX, No. 6, pp. 111–120.
- [14] D. Gee-Clough, M. Mcallister, G. Pearson and D. W. Evernden. « *The empirical prediction of tractor-implement field performance* ». Journal of

Terramechanics, 1978, Vol 15, No 2, pp 81 to 94. Pergamon Press Ltd. Printed in Great Britain. 1978.

- [15] H. Raheman, S.K. Jha. « *Wheel slip measurement in 2WD tractor* ». ScienceDirect. Journal of Terramechanics 44 (2007) 89–94. Agricultural and Food Engineering Department, Indian Institute of Technology, Kharagpur 721302, India, june 2006.
- [16] Rodney A. Rohrer. « *Investigation of Petroleum Use in Off-road Agricultural Machinery and Analysis of J1939 Controller Area Network (CAN) Data for Advanced Machinery Testing* ». Thesis. Presented to the Faculty of The Graduate College at the University of Nebraska In Partial Fulfillment of Requirements For the Degree of Master of Science. University of Nebraska – Lincoln, United States, 2017.
- [17] Lukasz Rymaniak, Piotr Lijewski, Michalina Kamińska, Paweł Fuć, Beata Kurc, Maciej Siedlecki, Tomasz Kalociński, Alex Jagielski. « *The role of real power output from farm tractor engines in determining their environmental performance in actual operating conditions* ». Elsevier. Computers and Electronics in Agriculture. Poznan University of Technology, Poland, 2020.
- [18] Rodney A. Rohrer, Joe D. Luck, Santosh K. Pitla, Roger M. Hoy. « *Evaluation of the accuracy of machine reported CAN data for engine torque and speed (j1939) data for engine torque and SPE* ». DigitalCommons@University of Nebraska - Lincoln, United States, 2018.
- [19] Samuel E. Marx, Joe D. Luck, Santosh Pitla, Roger M. Hoy. « *Comparing various hardware/software solutions and conversion methods for Controller Area Network (CAN) bus data collection* ». DigitalCommons@University of Nebraska - Lincoln, United States, 2016.
- [20] National Instrument Corp. « *The NI TDMS File Format* ». Consulté le 10 juinl 2021]. Consultable à l'adresse : <https://www.ni.com/en-us/support/documentation/supplemental/06/the-ni-tdms-file-format.html>.

- [21] Seong-Min Kim, Seung-Jae Park, Cheol-Soo Kim, Myeong-Ho Kim. « *Implementation of the communication model for ISO11783 standards based on AUTOSAR* ». An 2011 ASABE Annual International Meeting Presentation. Paper Number : 1111403. Dept. of Bioindustrial Machinery Engineering, Chonbuk National Univ. Jeonju, Korea, 2011.
- [22] William B. Vlcek. « *Implementing the CAN Calibration Protocol (CCP) in an SAE J1939 Application* ». Ascent Technologies, Inc. 525 Avis Drive, Suite 15 Ann Arbor, MI 48108 USA.
- [23] Patrik Soltic. « *On-Road Measurements, Test Bench Measurements and Emission Simulations for a Tractor-Semitrailer in Trans-Alpine Operation* ». EMPA Materials Science and Technology Laboratory for Internal Combustion Engines, Dübendorf, Switzerland, November 2004.
- [24] S. C. Borgelt, J. D. Harrison, K. A. Sudduth, S. J. Birrell. « *Evaluation of GPS for applications in precision agriculture* ». 1996 American Society of Agricultural Engineers.
- [25] James Bao-Yen Tsui. « *Fundamentals of Global Positioning System Receivers* ». A Wiley Interscience Publication, John Wiley & Sons, Inc. 2000. (Chapter 5, pages 73 - 105) Print ISBN 0-471-38154-3.
- [26] A. Suvinen, M. Saarilahti. « *Measuring the mobility parameters of forwarders using GPS and CAN bus techniques* ». ScienceDirect. Journal of Terramechanics 43 (2006) 237–252. University of Helsinki, P.O. Box 27 (Latokartanonkaari 7), FIN-00014, Finland, 2006.
- [27] Aradhya Biswas, Goutham Pilla, and Bheemarjuna Reddy Tamma. « *Microsegmenting : An approach for precise distance calculation for GPS based ITS applications* ». IEEE Recent Advances in Intelligent Computational Systems (RAICS). Indian Institute of Technology Hyderabad, India 2013.
- [28] Corentin Leroux. « *Géopositionnement en agriculture* ». Montpellier. Mise à jour le 3 juin 2020. [Consulté le 15 juin 2021]. Consultable

à l'adresse : <https://www.aspexit.com/geopositionnement-en-agriculture/>.

- [29] Andrew D. Townshend, Charles J. Worringham, and Ian B. Stewart. « *Assessment of Speed and Position during Human Locomotion Using Nondifferential GPS* ». Applied Sciences. Institute of Health and Biomedical Innovation and School of Human Movement Studies, Queensland University of Technology, Queensland, Australia, 2007.
- [30] Arijit Chowdhury, Tapas Chakravarty, P. Balamuralidhar. « *Estimating True Speed of Moving Vehicle using Smartphone-based GPS Measurement* ». 2014 IEEE International Conference on Systems, Man, and Cybernetics October 5-8, 2014, San Diego, CA, USA..
- [31] Prof. Nitin R.Chopde, Mr. Mangesh K. Nichat. « *Landmark Based Shortest Path Detection by Using A* and Haversine Formulal* ». International Journal of Innovative Research in Computer and Communication Engineering, India 2013.
- [32] L.H. Cao, J.N. Li, X.L. Liu, J.L. Yu, Z.M. Wu, F.G. Chen. « *Study on Diesel Engine Status Analysis and Fault Diagnosis Based on SAE J1939 Protocol* ». International Conference of Electrical, Automation and Mechanical Engineering (EAME 2015). Control Engineering Key Laboratory, Chongqing Communication Institute, Chongqing, China 2015.
- [33] Marco Bietresato, Aldo Calcante, Fabrizio Mazzetto. « *A neural network approach for indirectly estimating farm tractors engine performances* ». Fuel 143 (2015) 144–154. 0016-2361/O 2014 Elsevier Ltd.
- [34] Maohua Xiao , Weichen Wang, Kaixin Wang, Wei Zhang, and Hengtong Zhang. « *Fault Diagnosis of High-Power Tractor Engine Based on Competitive Multiswarm Cooperative Particle Swarm Optimizer Algorithm* ». Hindawi. Shock and Vibration Volume 2020, Article ID 8829257, 13 pages. College of Engineering, Nanjing Agriculture University, Nanjing 210031, China.

- [35] Corentin Leroux. « *Comment valider un modèle de prédiction ?* ». Montpellier. Mise à jour le 16 août 2018. [Consulté le 20 juillet 2021]. Consultable à l'adresse : <https://www.aspexit.com/comment-valider-un-modele-de-prediction/>.

- [36] Yousef Abbaspour-Gilandeh, Fatemeh Rahimi-Ajdadi, Ezzatollah Askari Asli-Ardeh, Vali Rasooli Sharabiani. « *Application of Artificial Neural Network for Predicting Fuel Consumption of Tractor* ». International Agricultural Engineering Conference, Bangkok, Thailand, 7 – 10 December, 2009.

- [37] Elton Gani, Chris Manzie. « *Indicated Torque Reconstruction from Instantaneous Engine Speed in a Six-cylinder SI Engine Using Support Vector Machines* ». 2005 SAE World Congress Detroit, Michigan.

- [38] Thorsten Joachims. « *SVMlight* ». Mise à jour le 14.08.2008. [Consulté le 5 août 2021]. Consultable à l'adresse : https://www.cs.cornell.edu/people/tj/svm_light/.

- [39] Lijuan Wang, Jeffrey Gonder, Eric Wood, and Adam Ragatz. « *The Accuracy and Correction of Fuel Consumption from Controller Area Network Broadcast* ». Copyright - 2017 SAE International.